

Cosmological gravity on all scales. Part II. Model independent modified gravity N -body simulations

Sankarshana Srinivasan,^a Daniel B. Thomas,^{b,a} Francesco Pace^{c,a}
and Richard Battye^a

^aJodrell Bank Centre for Astrophysics, School of Natural Sciences, University of Manchester,
Alan Turing Building, Oxford Road, Manchester, M13 9PL, U.K.

^bSchool of Physics and Astronomy, Queen Mary University of London,
G.O. Jones Building, 327 Mile End Road, London, E1 4NS, U.K.

^cDipartimento di Fisica ed Astronomia “Augusto Righi”,
Alma Mater Studiorum Università di Bologna,
Via Gobetti 93/2, I-40129 Bologna, Italy

E-mail: sankarshana.srinivasan@postgrad.manchester.ac.uk,
dan.b.thomas1@gmail.com, francesco.pace9@unibo.it,
richard.battye@manchester.ac.uk

Received March 16, 2021

Accepted May 19, 2021

Published June 8, 2021

Abstract. Model-independent constraints on modified gravity models hitherto exist mainly on linear scales [1]. A recently developed formalism presented a consistent parameterisation that is valid on all scales [2]. Using this approach, we perform model-independent modified gravity N -body simulations on all cosmological scales with a time-dependent μ . We present convergence tests of our simulations, and we examine how well existing fitting functions reproduce the non-linear matter power spectrum of the simulations. We find that although there is a significant variation in the accuracy of all of the fitting functions over the parameter space of our simulations, the ReACT [3] framework delivers the most consistent performance for the matter power spectrum. We comment on how this might be improved to the level required for future surveys such as *Euclid* and the Vera Rubin Telescope (LSST). We also show how to compute weak-lensing observables consistently from the simulated matter power spectra in our approach, and show that ReACT also performs best when fitting the weak-lensing observables. This paves the way for a full model-independent test of modified gravity using all of the data from such upcoming surveys.

Keywords: cosmological simulations, modified gravity, dark energy theory, weak gravitational lensing

ArXiv ePrint: [2103.05051](https://arxiv.org/abs/2103.05051)



Contents

1	Introduction	1
2	Modified gravity framework	2
3	Methodology	5
3.1	N -body simulations	5
3.2	Predicting the power spectrum on non-linear scales	8
4	Simulation phenomenology	13
5	Impact on weak-lensing observables	16
5.1	Computing weak-lensing observables in modified gravity	16
5.2	Further evaluation of fitting functions	20
6	Discussion and conclusion	21
A	Convergence tests	22
B	Binning in μ	23
C	Fitting functions	25
C.1	Cui et al.	25
C.2	Halo model	26
C.3	Halo model reaction	27

1 Introduction

In the current cosmological paradigm, dark matter (DM) and dark energy (DE), the so-called dark sector, are the dominant components of the energy density of the Universe, but their underlying physics is still unknown. In addition, assuming that General Relativity (GR) (plus the cosmological constant Λ) is the fundamental law of gravity is an extrapolation of many orders of magnitude (from the Solar System regime where GR is well-tested). Furthermore, the cosmological constant and its theoretical foundations are yet to be completely understood [4]. Dark matter has also hitherto eluded direct detection efforts [5, 6]. These problems are key parts of the theoretical motivation for testing gravity.

The modified gravity model space is vast [see [7, 8] for reviews of this topic]. As a result, in order to constrain significant volumes of this model space, one needs to formulate model-independent parameterisations that cover a large fraction of the model space.¹ There are many cosmological parameterisations of modified gravity [9–14], however the vast majority of these are developed in the “linear” regime, where perturbations (particularly density perturbations) are small. These have been constrained with multiple observational

¹Of course, a complete model independence is very difficult. For example, most modified gravity approaches assume a perturbed FLRW spacetime.

probes [15–21], but no deviation from Λ CDM has been detected universally. However, the restriction to linear scales significantly limits what these parameterisations can achieve.²

This restriction represents a significant challenge, since upcoming cosmological surveys such as *Euclid*,³ the Vera Rubin Observatory,⁴ the Nancy Roman Space Telescope⁵ and the Square Kilometre Array (SKA)⁶ will generate a multitude of data on non-linear cosmological scales. Until now, various studies of structure formation were carried out by performing N -body simulations in specific modified gravity models, such as $f(R)$ gravity [24] or Dvali-Gabadadze-Porrati (DGP) gravity [25]; for a review, see [26, 27]. While these studies are valuable in understanding the behaviour of the respective models in the non-linear regime, they do not allow one to rule out significant regions of the modified-gravity model space. Furthermore, these studies don’t allow us to perform a robust *null test* of GR, unlike model-independent approaches.

To this end, it is important to establish a model-independent approach towards testing modified gravity on all cosmological scales. Recent work [2] used the Post-Friedmann formalism [28] to design a rigorous approach to modified gravity that is valid on all cosmological scales. Modified gravity models are typically parameterised by two parameters in linear theory. The post-Friedmann approach extends these linear theory parameterisations in a manner that can be consistently simulated and interpreted on all scales. In this work we explore the non-linear phenomenology of this model-independent approach using N -body simulations. Crucially, a careful choice of the parameters ensures that we can run simulations characterised by a single parameter, and include the effect of the other in post-processing (see sections 2 and 5 for more details).

Understanding this behaviour, and being able to capture it using fitting functions is important if we are to use this approach to analyse the data from next generation surveys. These missions will generate a multitude of data, particularly on non-linear cosmological scales. Not only will they measure the matter power spectrum through galaxy clustering, but they will also measure cosmic shear: the apparent distortion of the shapes of galaxies due to the deflection of light by the distribution of mass along the line-of-sight. As such, being able to compute these weak-lensing observables from our simulations and any fitting functions is also of paramount importance.

The plan of this work is as follows. In section 2, we provide a brief review of the Post-Friedmann formalism and discuss some features of our approach and their impact on the simulations, followed by a description of our N -body pipeline in section 3 as well as a discussion on the performance of the different fitting-functions. We present the phenomenology of the simulations in section 4. In section 5, we compute weak-lensing observables in our approach from the simulations and examine the performance of the fitting functions in this context. We present our concluding remarks in section 6.

2 Modified gravity framework

Cosmological perturbation theory is typically done in the background of the FLRW metric with small perturbations. Perturbation theory enables us to impose model-independent

²Note that some recent works [22, 23] are valid on non-linear scales; we will return to these and contextualise them later in this work.

³<https://www.euclid-ec.org>.

⁴<https://www.lsst.org>.

⁵<https://roman.gsfc.nasa.gov>.

⁶<https://www.skatelescope.org>.

constraints on modified gravity models on linear scales [see for example [9] and references therein]. Meanwhile, a lot of progress has been made [29] in constraining GR on local solar-system scales by expanding the local Minkowski metric to varying powers of $1/c$, where c is the speed of light, with the aim of capturing local non-linear effects. Such an expansion is done assuming the Newtonian limit, i.e., the quasi-static approximation (where time derivatives in the metric perturbations are down-weighted). The Post-Friedmann formalism [28] generalises the Newtonian limit to weak-field cosmology, in a manner that applies to all scales. The FLRW metric is expanded with perturbations where $1/c^2$ and $1/c^3$ terms are considered leading order. The full metric is written as

$$g_{00} = - \left[1 - 2\frac{U_N}{c^2} + \frac{1}{c^4}(2U_N^2 - 4U_P) \right] + \mathcal{O}\left(\frac{1}{c^6}\right), \quad (2.1)$$

$$g_{0i} = -\frac{a}{c^3}B_i^N - \frac{a}{c^5}B_i^P + \mathcal{O}\left(\frac{1}{c^7}\right), \quad (2.2)$$

$$g_{ij} = a^2 \left[\left(1 + \frac{2V_N}{c^2} + \frac{1}{c^4}(2V_N^2 + 4V_P) \right) \delta_{ij} + \frac{1}{c^4}h_{ij} \right] + \mathcal{O}\left(\frac{1}{c^6}\right), \quad (2.3)$$

where the spatial Cartesian coordinates are as measured in an Eulerian system of reference. We operate in the Poisson gauge, one of the few gauges that is valid on both linear and non-linear cosmological scales [23]. The vector functions B_i^N and B_i^P are divergenceless and h_{ij} is a transverse and traceless tensor (see [28] for a detailed explanation of this expansion and the physical interpretation of each term). In this approach, the dominant constituents of the cosmological fluid are pressure-less dust and the cosmological constant. The Einstein equations and the covariant conservation of energy-momentum are used to obtain a closed system of equations.

Recently, the post-Friedmann formalism was studied in the context of validating a parameterisation of modified gravity models on both linear and non-linear scales [2]. This work uses the similarities between the linear-perturbative and Newtonian limits, and the lack of an intermediate regime where both limits fail, to create a set of “master” equations that can be applied on all cosmological scales. In this work, the so-called “re-summed potentials” [28]

$$\psi_P = -V_N - \frac{2}{c^2}V_P, \quad (2.4)$$

$$\phi_P = -U_N - \frac{2}{c^2}U_P, \quad (2.5)$$

are used to define the parameterisation for modified gravity models that we use for our simulations. As discussed in [2], the re-summed potentials, in combination with the fact that the weak-field approximation is valid on all cosmological scales, show that the same degrees of freedom and dynamics are operating on all cosmological scales. A parameterised Poisson equation that governs the gravitational dynamics of massive particles may be written as

$$\frac{1}{c^2}k^2\tilde{\phi}_P = -\frac{1}{c^2}4\pi a^2\bar{\rho}G_N\mu(a,k)\tilde{\Delta}, \quad (2.6)$$

$$\tilde{\psi}_P = \eta(a,k)\tilde{\phi}_P, \quad (2.7)$$

where $\bar{\rho}$ is the background density, $\tilde{\Delta} = \left(\tilde{\delta} - \frac{\dot{a}}{a} \frac{3}{c^2 k^2} i k_i \tilde{v}_i \right)$ is the gauge-invariant density contrast in Fourier space, FP: G_N Newton’s gravitational constant and $\mu(a,k)$ is an effective dimensionless Newton’s gravitational “constant” now promoted to a general function of space

and time. In Λ CDM μ is unity, but in modified gravity models this parameter may in principle be a complicated function of space and time. The work in [2] showed that these parameterised equations are valid on all cosmological scales, and encompass any modified gravity model as long as certain conditions are satisfied, such as the validity of the Newtonian limit of that theory on all non-linear scales. In particular, $f(R)$ [30] fits within this approach. Moreover, the works ([22, 31]) can be re-interpreted in this framework as specifying the functional forms of μ that occur for some commonly studied models with screening mechanisms.

A similar and promising extension of the linear theory parameters to a much wider range of scales was developed in [32], which applies to all models that fit into a Parameterised Post Newtonian (PPN) formalism [29]. The equivalent μ in this approach has a functional form such that it has scale dependent features on linear scales, but is purely time-dependent on non-linear scales, due to the restriction to theories that fit into the PPN framework. Another promising proposal to address parameterising modified gravity theories on non-linear scales was proposed in [22]. This theoretical setup maps specific modified gravity models to known screening mechanisms via detailed spherical collapse calculations. This formalism was recently implemented to run phenomenological N -body simulations of $f(R)$ gravity without having to solve for the additional dynamical scalar field [31]. As discussed above, this work can be re-interpreted as a description of the functional forms of μ in our approach for known theories with screening mechanisms. We note that our approach is less restrictive in theory space than either of the two approaches discussed here, however this comes at the cost of an increased difficulty to pinpoint the type of theory responsible if a deviation is found.

With all these points in mind, in this work we take the first step toward a fully model-independent treatment of modified gravity with N -body simulations. Various approaches for modelling μ are considered in [2] including, for example, using the known functional forms of μ on linear scales to transition from into the non-linear regime at k_{NL} (traditionally chosen to be the scale at which $k^3 P(k) = 1$). Such an approach would therefore have an implicit time-dependence, depending on when the transition to non-linear structure formation occurs for a given μ . For our simulations, we use the “maximally phenomenological pixels” approach as described in [2], in which we run N -body simulations where μ is a piece-wise constant function composed of independent bins or pixels in time and space which can be varied independently of each other. This allows us to be as agnostic as possible about the functional forms of the modifications to gravity. Note that this approach has been applied successfully in linear theory for modified gravity [1] and also generalised constraints on dark matter [33]. Since the evolution of matter is completely determined by the Poisson equation given in eq. (2.6), we can run the N -body simulations with a single modified gravity parameter μ , and only require the second parameter when computing the observables from the simulation outputs (see section 5 for a more detailed discussion of how our simulations can be used to probe both parameters). Since we restrict ourselves to sufficiently sub-horizon scales, the Newtonian approximation allows us to set $\tilde{\Delta} = \tilde{\delta}$, as is typical in Newtonian N -body simulations.

We will focus on models of modified gravity with a purely time-dependent μ . Whilst the full time- and scale-dependent analysis would employ similar methods and techniques outlined in this work, it requires a more involved alteration of the methods used to evolve the N -body simulation. We will present this in future work. More importantly, the purely time-dependent μ case is of considerable interest in its own right for several reasons. The time-dependent case is a bridge between linear and non-linear studies, because the linear growth factor remains purely time-dependent. This allows the non-linear scale dependent phenomenology due to an evolving strength of gravity to be clearly studied separately from

the additional complexity introduced even on linear scales when μ is scale dependent. This was shown in [34] for a general class of modified gravity models on linear scales. We will return to this point later in this work (see eq. (2.8) and the discussion following for further details on the utility of maintaining scale independence on linear scales). We leave combined scale- and time-dependent μ simulations to future work.

Previous studies of phenomenological N -body simulations were carried out in [35–39].⁷ As discussed in [2], the interpretation of these simulations is not always entirely clear, in terms of how they relate to linear parameterisations, and how they relate to the equations on non-linear scales in particular modified gravity theories. By working from eq. (2.6) and the derivation and framework of [2], we avoid these possible problems, although we note that some of these earlier simulations are justified *post-hoc* by [2]. The works [37, 38] restrict themselves to smaller regions of parameter space, while the studies [35, 36] are similar to ours in the sense that they concentrate on a purely time-dependent μ , although these simulations were run with a constant value for μ throughout the simulation. For this simplified setup, the authors in [35] derive a fitting function to calculate the power spectrum for arbitrary values of μ (see the appendix in [35] for details). In this work we explore the validity of the fitting function derived in [35], and other fitting functions in the literature, which attempt to reproduce the non-linear matter power spectrum $P(k)$ for modified gravity models.

Our simulations are the first to explore the phenomenology of μ having *different values in multiple redshift bins* through the simulation. This means that our method allows one to be sensitive to any time variation that μ may assume, irrespective of theoretical prejudices/biases.

As noted, an advantage of a purely time-dependent μ is that the linear growth factor $D = \delta/a$ is purely time-dependent, as in Λ CDM. We use this property in section 4 to separate whether phenomenology is sensitive to the era in which the modified growth occurred, and provide analytic checks of the simulations on large scales. It is also useful for comparing to different Λ CDM simulations, and when considering the “pseudo spectrum” in ReACT (see section 3). We solve for the linear growth factor $D(z)$ using

$$D'' + \left(2 + \frac{H'}{H}\right) D' - \frac{3}{2} \frac{\mu \Omega_m a^{-3}}{E(z)^2} D = 0, \quad (2.8)$$

where $'$ denotes a derivative with respect to the logarithm of the scale factor a and $E(z)$ is the dimensionless Hubble function.

3 Methodology

3.1 N -body simulations

In this section we present the details of the N -body simulations and code modifications, as well as the tests and consistency checks we performed. We use the publicly available N -body code GADGET-2 [43] as the basis for our N -body pipeline. GADGET-2 and modifications of it have been used extensively in the literature to run dark matter-only simulations of a wide variety [44–46]. We restrict ourselves to dark matter-only simulations in order to understand modified gravity phenomenology. We present a summary of our numerical pipeline in the flowchart presented in figure 1. We will now discuss the main branch of this pipeline in detail.

⁷As noted in [2], scalar field models typically arrive at a modified Poisson equation as well, so N -body simulations that seek to treat scalar field models quite generally (e.g. [40–42]) will be naturally similar to the code modifications we discuss here, although the justification and interpretation is different.

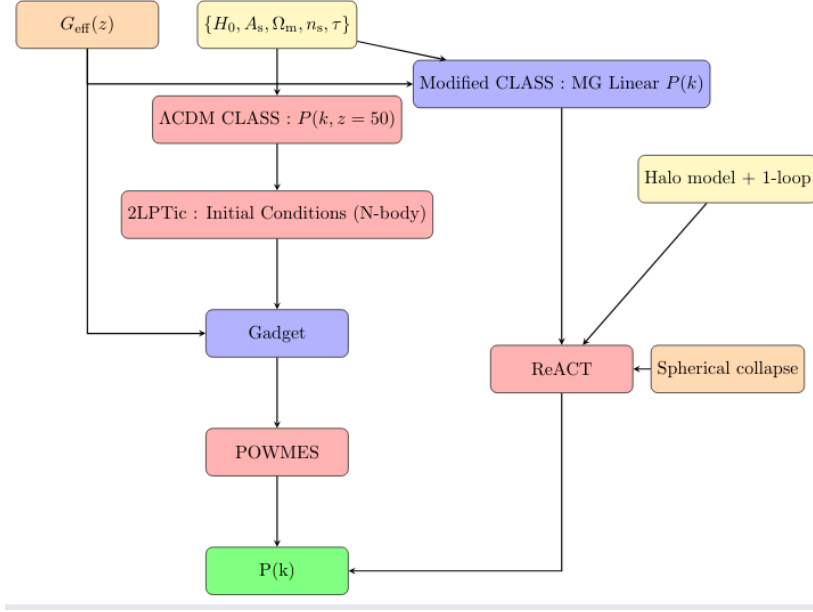


Figure 1. Flowchart of the numerical pipeline that we implemented. We establish initial conditions at $z = 50$, which are then evolved up to $z = 0$ using **GADGET-2**. Blue panels represent codes that we have modified, while red panels are codes that are used in their publicly available form. The modified gravity parameters are in orange, while Λ CDM parameters are in cream. The central branch represents our numerical simulation pipeline, while the right hand branch denotes our most successful pipeline for fitting the simulations (see section 3.2 for more details).

GADGET-2 employs a TreePM algorithm that splits the force computation into a long-range and short-range force. We modify both the long-range and the short-range force in the algorithm. We follow a similar procedure as in [47], where we have a modified Poisson equation with a time-dependent μ parameter, which is determined as a function of the simulation redshift. We now describe the time evolution of μ within our simulations.

For the phenomenological pixels that we are considering, the μ function is split into a set of piece-wise constant redshift bins, i.e., each with a constant value for μ . In order to explore the phenomenology of this approach as well as make contact with earlier work [35], we run a suite of simulations with a simple “1-2-4” hierarchy of pixels. In other words, we split the redshift interval $0 \leq z \leq 50$ into 1, 2, and 4 redshift bins of equal Λ CDM growth. In order to isolate as far as possible the non-linear behaviour (of the matter power spectrum) from the linear theory behaviour, we choose the values of μ in each bin such that the linear growth factor at $z = 0$ is equal when each bin is “switched on” exclusively, with $\mu = 1$ in all the other bins (see table 1 and figure 2 for the redshift bins and the respective μ values that correspond to each of them). The widths of each bin are chosen such that, for our cosmological parameters, there is equal growth in each bin in a Λ CDM universe. See appendix B for full details of this procedure. This allows an additional consistency check for our simulations, since the matter power spectrum at $z = 0$ for all our simulations should differ from the standard Λ CDM result by a factor $[D(z)]^2$.

We use 2nd order Lagrangian perturbation theory [48] to generate our initial conditions at $z = 50$ for all our simulations, via the **2LPTic** code [49]. We only consider modifications to gravity from the starting redshift of the simulations. This allows us to use the same initial

Number of bins	Redshift	μ	
1	0-50	1.044	0.956
2	0-7.0	1.100	0.900
	7.0-50	1.080	0.920
4	0-2.1	1.256	0.746
	2.1-7.0	1.167	0.837
	7.0-19.2	1.162	0.842
	19.2-50.0	1.161	0.843

Table 1. The redshift bins of equal Λ CDM growth according to the procedure described in the text, where we have the same linear growth at $z = 0$ when one of the bins is “switched on” for one, two and four bin(s). The μ values in each bin such that $P(k)$ at $z = 0$ when only that bin is turned on, is identical to $P(k)$ at $z = 0$ in the case when only the reference bin is turned on.

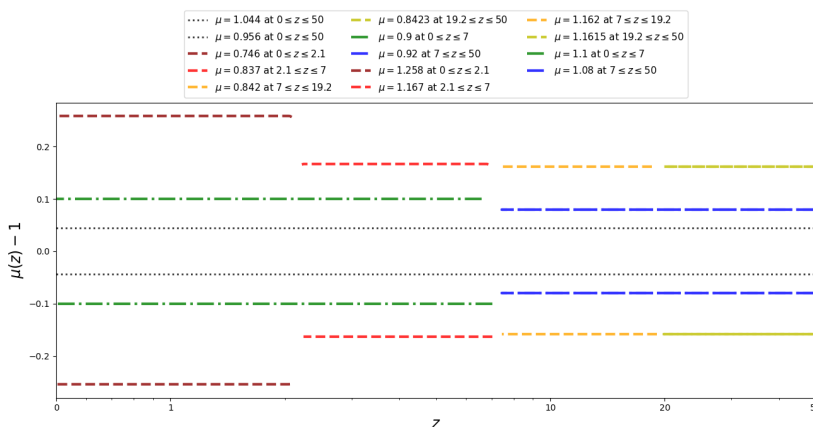


Figure 2. We illustrate the redshift evolution of μ in our simulations. Dotted lines represent the case where μ is constant through the simulation (one bin), dashed lines represent the case where there are two bins in redshift, while dot-dashed lines represent the case where there are four bins in redshift.

conditions (and therefore realisations) for modified gravity and Λ CDM, allowing us to focus on the phenomenology due to the modified gravitational evolution. We take advantage of these identical initial conditions as this allows us to present our results in terms of ratios of matter power spectra [50] (see appendix B for further details). Taking ratios of matter power spectra ensures that we cancel out any realisation dependent effects that individual models, i.e., μ values and redshift bins are insensitive to, up to a wavenumber $k \sim 5 h \text{ Mpc}^{-1}$ that was validated by convergence tests (see appendix A for more details). For all our simulations, we use a Λ CDM initial power spectrum created using the CLASS code [51] at $z = 50$, with the following parameters $\Omega_m = 0.315$, $\Omega_\Lambda = 0.685$, $h = 0.674$ and $\sigma_8 = 0.811$.

We also run a separate set of simulations with modified initial conditions such that we obtain Λ CDM matter power spectra with the same linear growth at $z = 0$ as our modified gravity simulations. These simulations have a modified value of $\sigma_8 = 0.883$ and $\sigma_8 = 0.742$, for the simulations with $\mu > 1$ and $\mu < 1$, respectively. This allows us to match the linear growth at $z = 0$ of all our simulations with the corresponding re-scaled Λ CDM simulation. This is important as the ReACT fitting function that we explore later in section 3.2 predicts

exactly the ratio of matter power spectra in modified gravity w.r.t. to Λ CDM spectra with matched linear growth at the same redshift. We refer to these Λ CDM spectra as ‘pseudo’ spectra. These simulations have the same initial seed as our standard Λ CDM simulations.

To measure the matter power spectrum from the output of our simulations, we use the publicly available `powmes` code [52, 53]. For notational convenience, we define the ratio $S(k) = P(k)/P_{\Lambda\text{CDM}}(k)$ to be the ratio of the matter power spectrum measured from our simulations to the Λ CDM power spectrum and $R(k) = P(k)/P_{\text{pseudo}}(k)$ to be the ratio of the modified gravity matter power spectrum from the simulations to the Λ CDM power spectrum with the same linear growth factor.

To test the convergence of our simulations, we run them increasing the particle numbers for an identical box size. We then plot the ratio of the matter power spectra from each simulation to the power spectrum from the one with highest resolution, which we choose to be our reference. Our reference simulation contains 1024^3 particles in a comoving box of side $250 h^{-1} \text{ Mpc}$. We demonstrate convergence in our Λ CDM simulations by showing that the power spectrum from the one with the second-highest resolution (512^3 particles) agrees with that of highest resolution up to $k = 4.8 h \text{ Mpc}^{-1}$, shown by the vertical green line in figure 11 (see appendix A for further details). We use a gravitational softening length equal to 2% of the inter-particle-distance, and a comoving mesh with the same number of grid points and particles.

We ensure that μ is a well-defined, continuous function of redshift by smoothing at the bin edges. In order to quantify the effect of such a smoothing procedure on the power spectrum, we first worked with the `CLASS` code, before implementing an identical algorithm in `GADGET-2`. Based on these results, the same smoothing procedure was implemented in `GADGET-2` (see appendix B for details). These results indicate that there is very little difference in the simulation results, regardless of whether μ is a smooth, continuous function of redshift or not, for our chosen parameters.

Note that we have primarily discussed the central branch of the flowchart in figure 1 in this section; we now turn to the right hand side branch of the pipeline in section 3.2.

3.2 Predicting the power spectrum on non-linear scales

N -body simulations are too computationally expensive to be used to forecast and perform data analyses for upcoming surveys. Thus in this section we test the performance of various theoretical formulations that attempt to predict the full non-linear matter power spectrum. We do this by comparing their respective accuracy in the reproduction of the ratio of the modified gravity matter power spectra with respect to the Λ CDM matter power spectrum as measured from our simulations. This comparison of ratios allows us to remain independent of realisation-dependent effects.

The results in [35] provide hints that knowledge of the linear power spectrum is insufficient to completely calculate the matter power spectrum on non-linear scales. This has important consequences for the fitting functions. In section 4 we show explicitly that in our framework simulations with identical linear matter power spectra can show significant differences in their non-linear power spectra.

We consider the following formalisms as candidates for predicting the non-linear matter power spectrum:

- case 1 — the fitting function obtained in [35];

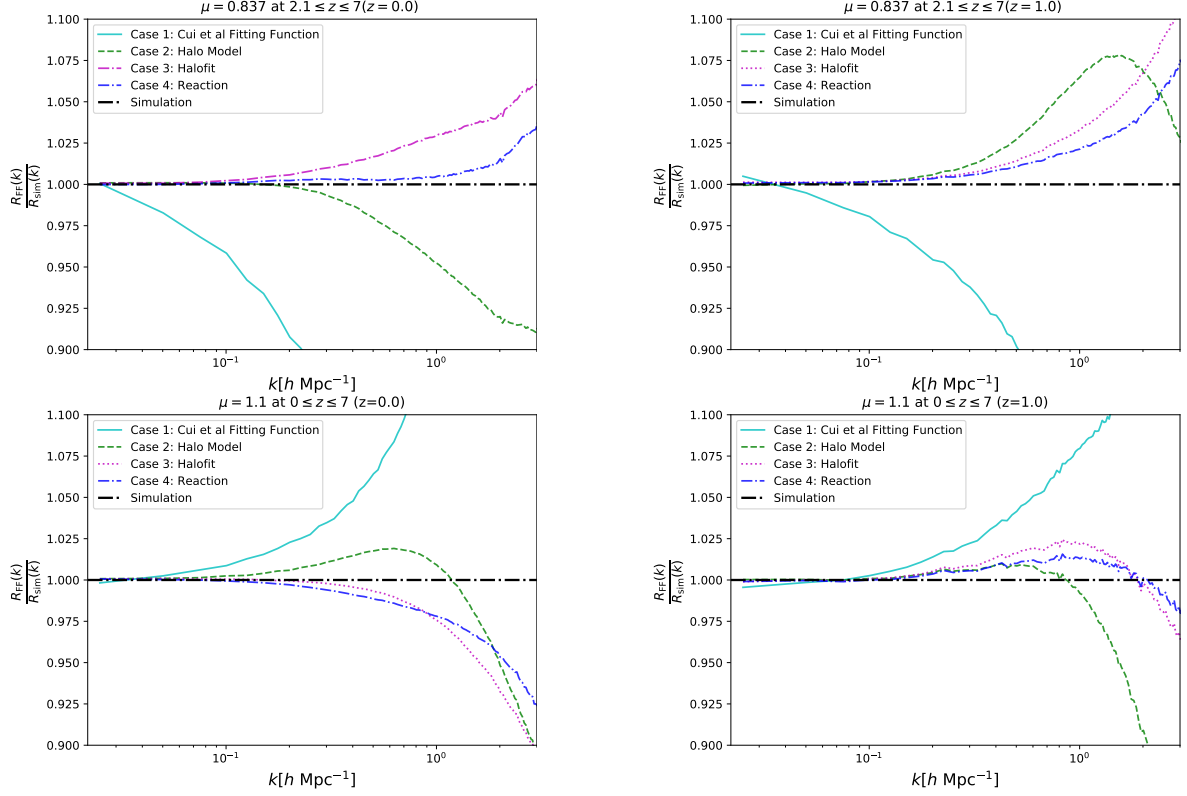


Figure 3. The ratio of the quantity $R(k)$ as predicted by the various fitting functions (as indicated by the superscript ‘FF’) with respect to the same quantity computed from the simulations (as indicated by the superscript ‘sim’). It is clear that the cyan line from the fitting function of [35] fails for all cases. The green dashed curve is the simple application of the halo model as shown in eqs. (C.2) and (C.3). It is important to note that the `halofit` prediction (magenta dotted) is actually identical for all the simulations, since they all have identical linear power spectra. In all the panels, we can clearly see the `ReACT` curve (blue dot-dashed) is consistently within 5% of the black dot-dashed line which indicates our reference, the simulations.

- case 2 — the standard halo model as in eqs. (C.2), (C.3) with a Sheth-Tormen mass function [54, 55];
- case 3 — the `halofit` fitting procedure [56];
- case 4 — the halo model reaction formalism [3] (`ReACT`).

We will now briefly describe each case.

To our knowledge, the only fitting function in the literature based on an arbitrary μ is that presented in [35]. As noted earlier, this fitting function was estimated assuming a constant μ through the simulations, which is not the case in our simulation (see appendix C for more details). Nevertheless, we investigate the performance of this fitting function in the context of our simulation results.

As suggested in [35], the halo model provides a potential explanation of the results obtained in the simulations. We implement the standard halo model to predict the matter power spectrum [see appendix C]. We also consider the `halofit` fitting procedure [56–58] that is commonly used for a Λ CDM Universe.

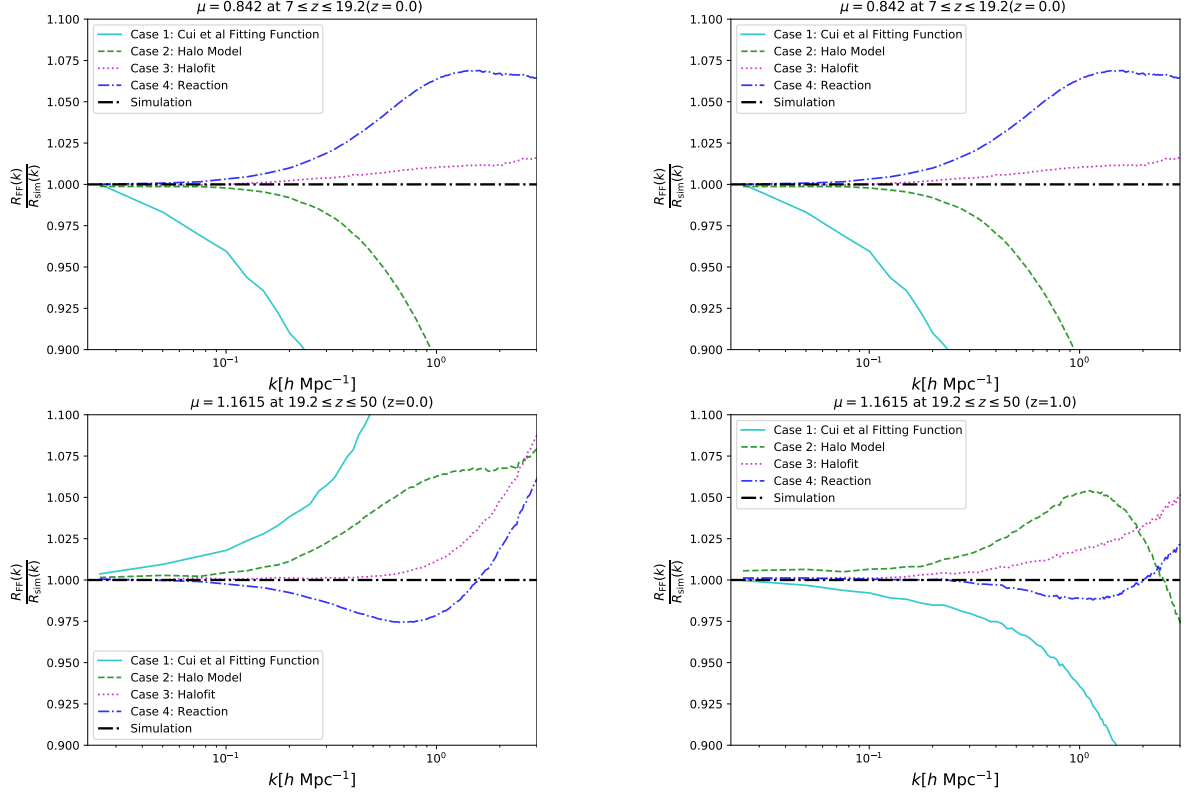


Figure 4. As in figure 3, but where μ has a different value from GR in one of the earlier redshift bins.

However, it is well known that the basic implementation of the halo model (the simple addition of the 2-halo and 1-halo terms, or equivalently, the addition of the 1-halo term to the linear matter power spectrum) is inaccurate on quasi-linear scales due to halo correlation effects that require more detailed modelling [59]. To circumvent the problems associated with the halo model on quasi-linear scales, the authors in [3, 60] modified the halo model. This can be seen in the following equation

$$P_{\text{mg}}(z, k) = \mathcal{R}(z, k) P_{\text{pseudo}}(z, k), \quad (3.1)$$

where $\mathcal{R}(z, k)$ is the so-called halo model reaction and P_{pseudo} is the Λ CDM ‘pseudo’ power spectrum with identical linear growth to P_{mg} . The reason for the use of the pseudo spectrum is that since it has identical linear clustering with respect to the modified gravity spectrum, the inaccuracy in the transition from linear to non-linear scales is slightly alleviated (see appendix C for a more detailed explanation of the various fitting functions).

The authors also invoke 1-loop perturbation theory corrections to quasi-linear scales for the halo model (see appendix C for more details). We note that the authors in [3] modify the Λ CDM concentration-mass relation for the dark energy models under consideration (see appendix C for more details and further discussion of the theoretical formulation of ReACT. See also figure 6 where we consider a further modification of the ReACT formalism).

In figures 3 and 4, we show the ratio of the quantity $R(k)$ (i.e., the ratio of the modified gravity matter power spectrum to the pseudo Λ CDM matter power spectrum) as predicted by the fitting functions with respect to the same quantity computed by the simulations. We

have chosen a subset of our simulations that is representative of the performance of the different fitting functions we consider. Clearly, we see that the fitting function in [35] fails to predict the correct non-linear behaviour when μ is not a constant value throughout the simulation. As expected, the **halofit** prediction is identical for all the simulations, since they all have identical linear growths. In other words, the simulations where the **halofit** performs better than **ReACT** are the cases where modified gravity parameters coincidentally result in a matter power spectrum that is already very close to that of Λ CDM. We show that on average, this accidental success is not typically replicated in the general case. We see that while the **ReACT** formalism doesn't always predict the non-linear trend perfectly, it appears to be consistently within 5% of the simulations. Since we are only interested in understanding the modified gravity phenomenology, we do not vary Λ CDM parameters in this work. We leave the full validation of **ReACT** across Λ CDM and modified gravity parameter space to future work.

We use the following least-square statistic to quantify the agreement between the various fitting functions with our simulations

$$\chi^2 = \frac{1}{N} \sum_{k=0.02 h \text{ Mpc}^{-1}}^{k=k_{\text{cut}}} [R_{\text{sim}}(k) - R_{\text{FF}}(k)]^2, \quad (3.2)$$

where $N = 160$, with the sampling in Fourier space being approximately logarithmic, as in **powmes** and k_{cut} is a wavenumber cut-off that we choose. The subscript 'sim' implies the ratio as measured in our simulations and the subscript 'FF' implies the ratio as predicted by the various fitting functions. We employ a cut-off at $k_{\text{cut}} = 3 h \text{ Mpc}^{-1}$, which is the theoretical threshold up to which the **ReACT** formalism is supposed to be accurate [3]. Note that this scale is within the region of validity of our simulations [see appendix A], which is $k_{\text{cut}} = 5 h \text{ Mpc}^{-1}$. We show the χ^2 statistic with both wavenumber cut-offs in figure 5. This indicates that the **ReACT** formalism provides the best fit to our simulations, as evidenced by the fact the red line is largest in the majority of the cases presented. Note that since we plot the negative logarithm, a larger bar corresponds to a better fit. We also note that due to the failure of the Cui et al. [35] (see eq. (C.1) in appendix C) even at small k to capture the results of our simulations, we don't include their fitting function in this figure.

We also show the wavenumber k_{fail} at which the quantity $R_{\text{FF}}(k)/R_{\text{sim}}(k)$ deviates from unity at the level of 3 and 5 %, respectively in the right panel of figure 5. Note that this probes the validity in k -space of each fitting function, while the χ^2 statistic is an indicator of the accuracy of each fitting function within this region of validity. The combination of k_{fail} and χ^2 throws light on the applicability of a fitting function to future forecasts.

In [35], the authors claim that in order to accurately reproduce the power spectra in simulations using the halo model, one needs to take into account that any modification to μ results in a change in the rate of structure formation, i.e., a change in the concentration parameter associated to haloes. Increasing μ causes haloes to form earlier, leading to lower concentrations in comparison to Λ CDM and vice-versa for decreasing μ . It is also worth noting that in order to accurately reproduce power spectra in $f(R)$ and DGP models, the authors in [3] modify the Λ CDM concentration-mass relationship. Therefore, we would expect that in order to achieve precision in forecasting phenomenological modified gravity, one would need to depart from the concentration-mass relationship observed in Λ CDM.

In anticipation of this, we carry out a simple modification of the **ReACT** formalism in which we rescale the Λ CDM concentration parameters for all halo masses. This re-scaling is given by $c_{\text{vir}}^{\text{MG}} = c_{\text{vir}}^{\Lambda\text{CDM}}/A$, where A is a function that would conceivably depend on μ ,

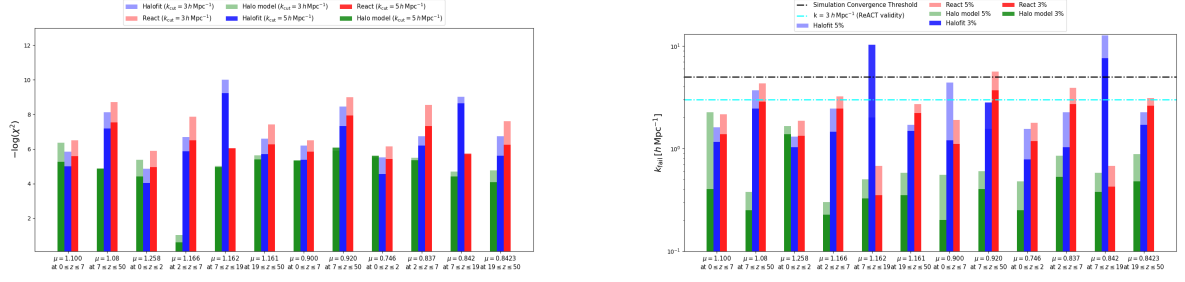


Figure 5. In the left panel, we show the values of the least square statistic for the different fitting functions as defined in eq. (3.2). We obtained these values by employing $k_{\text{cut}} = 3 h \text{ Mpc}^{-1}$ and $k_{\text{cut}} = 5 h \text{ Mpc}^{-1}$, which represent the wave number up to which the ReACT formalism is designed to operate and the wave number up to which we have converged results from our simulations, respectively. Clearly, the red bar is larger than the other bars in the majority of the simulations which implies that the ReACT formalism is typically the best fit over the range of scales (note that since we plot the negative logarithm, a larger bar is a better fit). In the right panel, we plot the wavenumber at which the quantity $R(k) - 1$ is at 3% and 5%, respectively. The results are consistent with the left panel, with the red bar being the largest for the majority of the simulations. This shows that the ReACT formalism allows one to probe deeper into the non-linear regime.

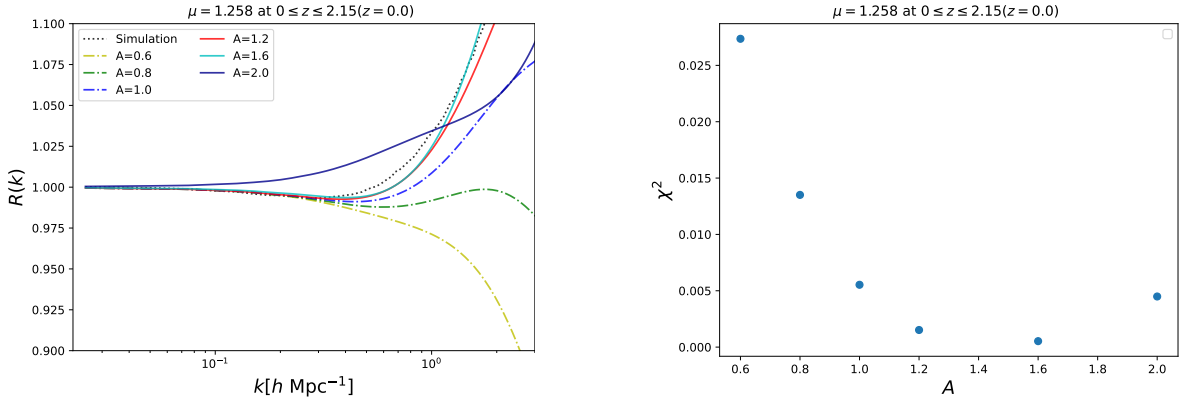


Figure 6. We display the ratio of the power spectra with respect to one of our simulations for several values of the parameter A in the left panel. The solid cyan line (which corresponds to $A = 1.6$) is the best fit, with the smallest χ^2 parameter, as can be seen by the plot on the right panel. While this behaviour was hinted at in various works in the literature [35], to our knowledge this is the first time that such an analysis has been carried out for the general model-independent case time-varying μ . We leave the analysis of varying A as a function of μ and the resulting rigorous validation of this modification to future work.

the redshift at which it is modified and the duration in redshift over which it is modified. In figure 6, we demonstrate that for different values of A , the non-linear tail of the power spectrum is significantly affected. Indeed, we find a better fit for $A = 1.6$, as can be seen in the right panel of figure 6. We note that this is only a preliminary examination of this effect and requires rigorous validation in future work, nevertheless we find similar qualitative behaviour to other studies.

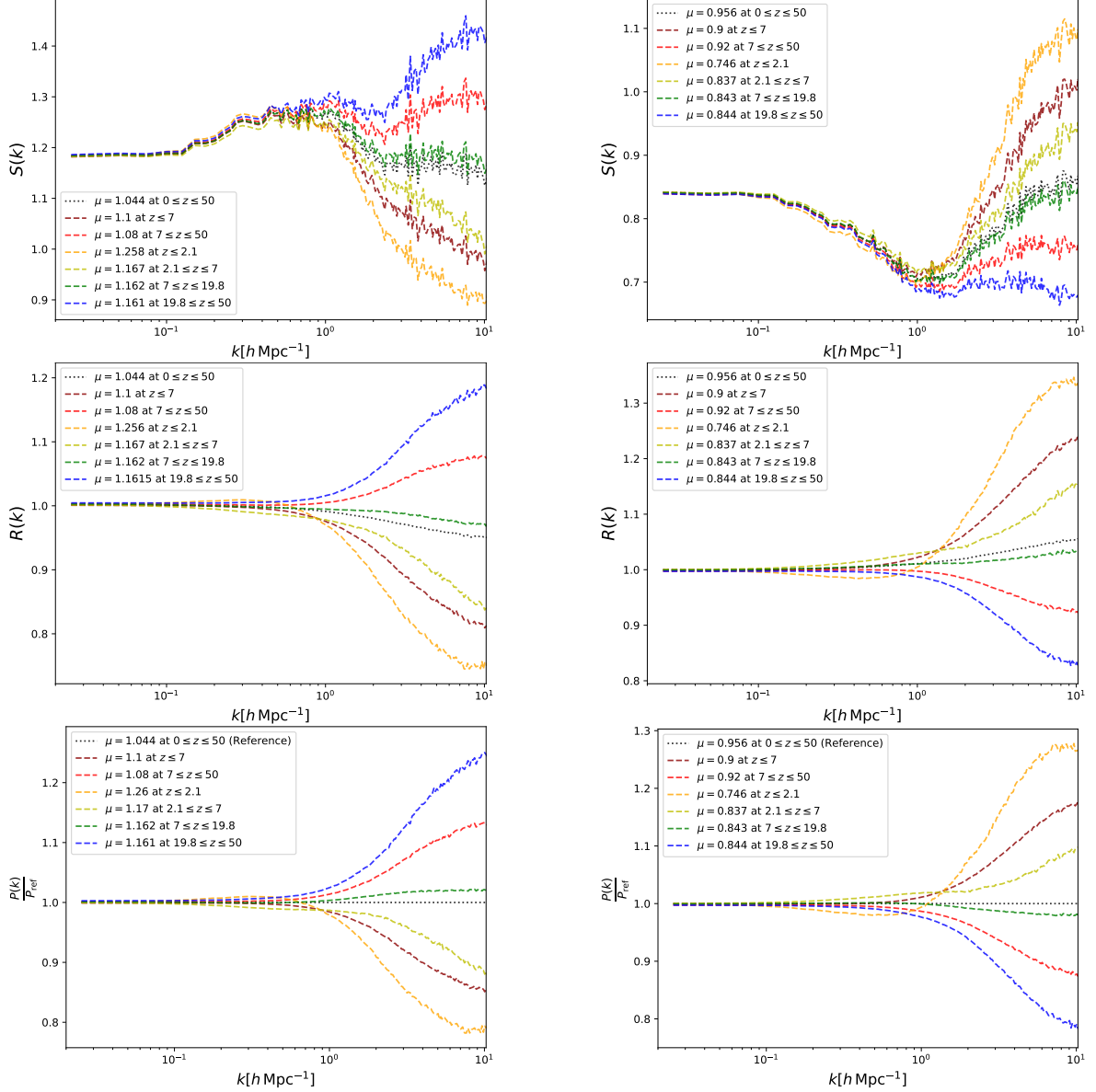


Figure 7. *Top row:* the ratio of the power spectra from all the simulations to the Λ CDM power spectrum at $z = 0$. Note that we see a quasi-linear bump in these due to the mismatch in linear growth, which is absent in the other rows. We see the multi-dimensionality of the parameter space in display here since we have 6 different unique non-linear power spectra for the same linear power spectrum as the case where μ is constant throughout the simulation as studied in [35]. *Middle row:* the quantity $R(k)$, the ratio of the matter power spectra from the different simulations to the pseudo Λ CDM power spectrum with the same linear growth. *Lower row:* ratio of power spectra at $z = 0$ from all the simulations to the power spectrum at $z = 0$ from the reference simulation with a constant μ throughout, with redshift bins according to table 1.

4 Simulation phenomenology

In this section, we discuss the phenomenology from our full suite of simulations with the fourteen different forms of $\mu(z)$. Our goal is to understand the variation in the non-linear

matter power spectrum as a function of $\mu(z)$. In the top row of figure 7, we present the ratio $S(k)$ (at redshift zero) of the power spectra from all the simulations to the Λ CDM spectrum with the same initial conditions. We note that while the ratio relative to the pseudo-spectrum $R(k)$ is the most appropriate ratio for the fitting functions in the previous section, here we focus on $S(k)$ because it throws light on the different structure formation histories of the simulations with respect to Λ CDM. By design, the lines in these plots are all equal on linear scales (for small k). We observe two distinct features in $S(k)$, present on quasi-linear scales and non-linear (large k) scales respectively. Firstly, we consistently see a rise (or dip, depending on whether $\mu > 1$ or $\mu < 1$) in power across all simulations on quasi-linear scales at the level of $\sim 20\%$ (at $0.1 h \text{ Mpc}^{-1} \leq k \leq 1 h \text{ Mpc}^{-1}$). This is followed by a “split” behaviour on small (non-linear) scales, where the power depends on the range of redshifts over which the modified gravity effects were switched on, and the simulations with $\mu \neq 1$ at large redshifts exhibit the opposite behaviour to the case where it is switched on at small redshifts. This is due to varying halo formation times and inter-halo clustering, as we will discuss later in this section.

The quasi-linear behaviour may be attributed to the difference in growth histories between Λ CDM and the different simulations, since the effect of a different growth rate is more pronounced on quasi-linear scales compared to linear scales. We illustrate this using the middle panels of figure 7, where we compute the ratio of the matter power spectrum in our simulations to the pseudo power spectrum. This quasi-linear feature is significantly diminished, since the pseudo power spectrum has the same linear growth at $z = 0$, so by construction the increased effects of the changed growth are mostly removed. This is why the pseudo power spectrum is a good comparison point when examining fitting functions. The same quasi-linear behaviour is seen in the bottom panel, where we compute the ratio of the matter power spectra relative to a reference simulation where μ is held constant in redshift.

We now turn to the non-linear split behaviour that we see in all of the rows. On small scales the lines diverge, depending on the redshift range during which $\mu \neq 1$ in each simulation. In particular, the simulations where $\mu \neq 1$ at earlier times, continues the trend expected from linear scales, that $\mu > 1$ increases the power and vice versa for $\mu < 1$. However, varying μ at low redshifts introduces the opposite effect to linear theory. For example, $\mu > 1$ at late times actually leads to a lack of power on non-linear scales. This results in “crossing points” with respect to Λ CDM, i.e. scales below which the power spectrum is actually reduced, despite equal initial conditions and $\mu > 1$. This implies that the state of clustering when μ is modified, in terms of how much structure has already formed and on what scales, is important for understanding the final clustering spectrum on small scales at redshift zero. Since we obtain intrinsically different shapes for the power spectra for different values of μ with identical linear growths, the discussion on concentration factors in [35] is insufficient to explain the physics of these simulations. These plots clearly show that any fitting procedures that predict the non-linear matter power spectrum purely from the linear spectrum will fail to capture non-linear effects accurately.

We can examine this small-scale issue in more detail by looking at the redshift evolution of the matter power spectrum in our simulations. In figure 8, we show the ratio of the power spectra from the simulations to the Λ CDM spectrum, at all of our output redshifts. There is a clear transfer of power over time from small scales to large scales. This is demonstrated in the top row where the peak/trough (depending on whether $\mu > 1$ or $\mu < 1$) in the ratio of the power spectra shifts to the left, i.e., to larger scales as one steps forward in redshift (from high redshift to low redshift). In the bottom panels, this transfer of power persists, although we

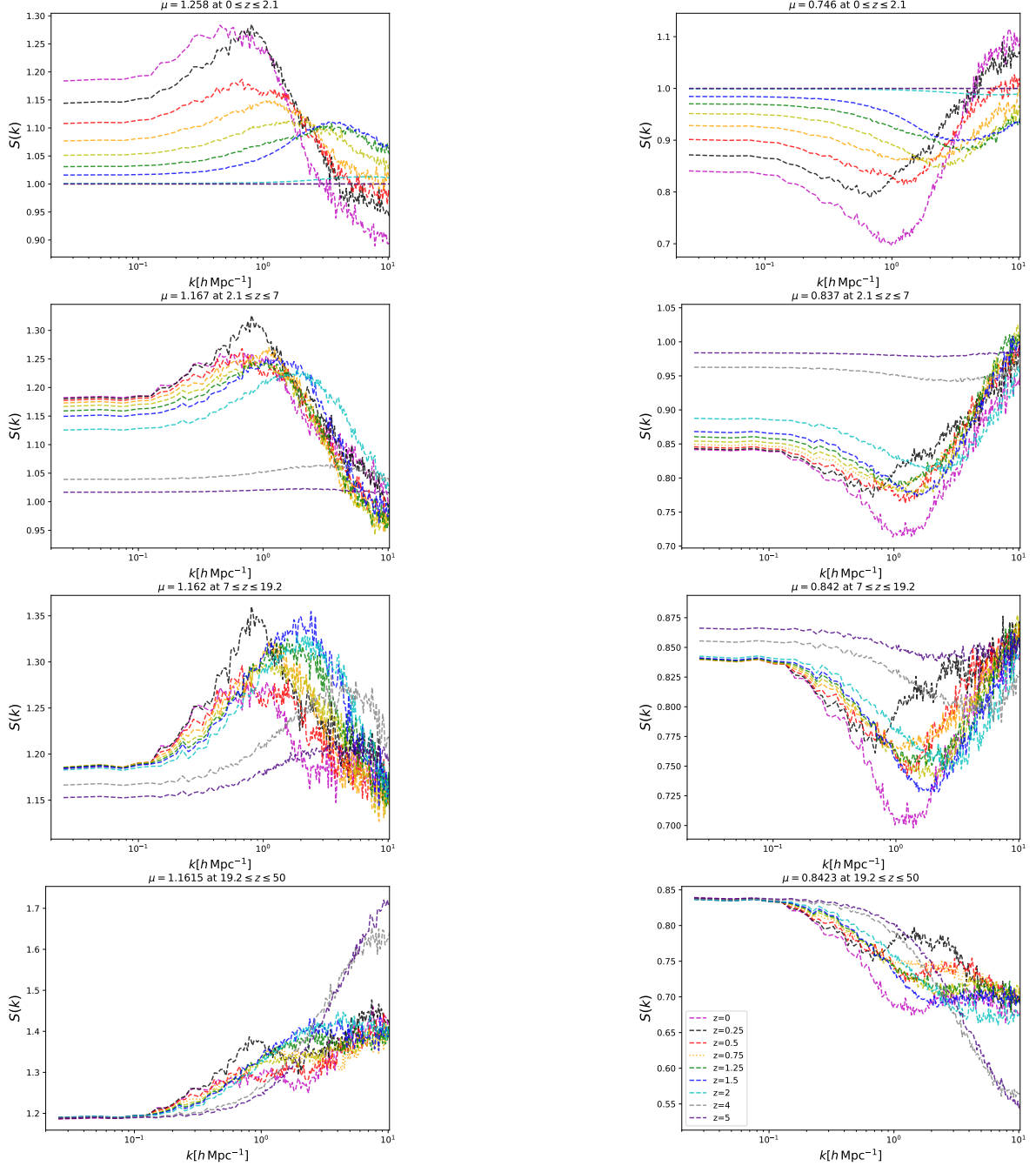


Figure 8. The ratio of the power spectra at all the redshifts output from our simulations relative to the Λ CDM power spectrum at the same respective redshifts, for the case where we have 4 bins in μ but the same linear growth at $z = 0$. These figures highlight the hierarchical nature of structure formation and its varying influence on the power spectrum depending on whether μ is switched on early in the simulation or later. In the former case, we have an early modification of the non-linear power spectrum, the imprint of which can be noticed at $z = 0$, while in the latter case, the late switching on of μ , introduces interaction between already formed structures resulting in non-intuitive shapes for the matter power spectrum.

see it manifested as phenomenon that ‘evens out’ the power over large and small scales. This transfer of power indicates the formation of larger structure in the universe causing power to be deposited on larger scales. Comparing the position and height of this peak/trough provides information on the rate of structure formation and the gravitational interaction at different epochs. On comparing the behaviour of the magenta, black and red lines in the top row with the bottom row, we also see the reinforcement of our earlier result, that there is an excess of power in the non-linear regime when $\mu > 1$ at high redshift, and vice versa at low redshift (this behaviour is mirrored when $\mu < 1$). We now consider these two cases in more detail.

Structure formation in the Universe is hierarchical, with smaller haloes being formed before larger ones. Therefore, modifying μ at early times impacts the power spectrum on the smallest scales as demonstrated in the bottom row of figure 8 where there is a substantial peak in $S(k)$ on small scales at high redshift $z = 4$ and $z = 5$. This is due to the fact that only the smallest haloes were in the process of formation. Effectively, the increase in the rate of structure formation at early times causes smaller haloes to be formed earlier. The formation of larger structure only takes place when growth reverts to the Λ CDM rate. As one steps forward in redshift, one sees that this small scale peak is diminished, but not totally destroyed. As larger haloes start to form, mergers begin to take place leading to a gradual shift in power towards larger scales. However, there is still an excess of power at $z = 0$ on small scales, as a signature of the enhanced structure formation at earlier times.

Modifying μ at late times introduces additional complexity, since at late times, larger structures have started to form. Therefore, increasing/decreasing the strength of gravity massively impacts the interaction between larger haloes, i.e., halo power spectrum as seen by the aforementioned shift in the peak of the power spectrum to the left in the top row of figure 8. Physically, this means that the interaction between haloes depletes power on smaller scales and deposits power on intermediate/large scales. Therefore, there is a transfer of power from smaller scales to larger scales (and of course the opposite happens when $\mu < 1$). This further validates the behaviour of the orange, dark red and yellow lines in figure 7, where the small scale behaviour is opposite to that on linear scales. We note that the middle rows in figure 8 contain a mix of the two limiting cases described in the above paragraphs.

5 Impact on weak-lensing observables

5.1 Computing weak-lensing observables in modified gravity

We now discuss the impact of varying μ in different redshift bins on the observables. In this section we compute the weak-lensing convergence power spectrum from the output of our simulations, and examine the impact of having predictions that are not restricted to linear scales. Furthermore, we also discuss how one can probe our two-parameter family approach to modified gravity from the output of our simulations.

The weak-lensing convergence power spectrum can be computed from the matter power spectrum and the modified gravity parameters from section 2 using the following expression

$$P_{\kappa}(\ell) = \frac{9H_0^4\Omega_m^2}{4c^4} \int_0^{\chi_{\max}} d\chi \frac{1}{a^2(\chi)} g^2(z) \frac{\mu^2(1+\eta)^2}{4} P_{\delta}(\ell/\chi), \quad (5.1)$$

where P_{δ} is the matter power spectrum, η is the second modified gravity parameter that affects the photon geodesics and χ is the comoving angular diameter distance to the source along the line of sight. This equation is derived using the standard procedure of solving for

null geodesics in the perturbed FLRW metric, in order to express the convergence in terms of the metric potentials, but we then replace the potentials by substituting the modified Poisson equation in eq. (2.6). We stress that in order to compute the matter power spectrum, one requires no knowledge of η . Therefore, we need only model μ in our N -body simulations. We will return to this point shortly. In practice, the above integral is truncated at the distance χ_{max} corresponding to the maximum source redshift the survey is sensitive to. In this work, we concentrate on the auto-power spectra. The additional complications involved in the analysis of the cross-power spectra requires a full statistical forecast, which is beyond the scope of this work.

The function $g(z)$ is a filter function that depends on the redshift distribution of the background galaxies and is typically written as

$$g_i(z) = \int_z^\infty dz' \left(1 - \frac{\chi(z)}{\chi(z')} \right) n_i(z'). \quad (5.2)$$

We use the standard expression for the source galaxy redshift distribution given by [61]

$$n(z) \propto \left(\frac{z}{z_0} \right)^\alpha \exp \left[- \left(\frac{z}{z_0} \right)^\beta \right]. \quad (5.3)$$

We adopt a Euclid-like binning of the source number density into 10 equi-populated bins according to eq. (5.3) with $z_0 = 0.9/\sqrt{2}$, $\alpha = 2$ and $\beta = 3/2$, where we have assumed an average source density $\bar{n}_g = 30 \text{ arcmin}^{-2}$ [62–66]. Due to the fact that the ReACT formalism is only valid at $z < 2.5$, we concentrate only on the first tomographic bin, i.e., $0 \leq z \leq 0.4$. The error bars are obtained by computing the following [66]

$$\delta C_\ell^{ij} = \sqrt{\frac{2}{f_{\text{sky}}(2\ell + 1)}} \left(C_\ell^{ij} + \frac{\sigma_\epsilon^2}{\bar{n}_i} \right), \quad (5.4)$$

where $f_{\text{sky}} = 0.7$ is the fractional sky coverage, $\sigma_\epsilon = 0.21$ is the variance of the observed ellipticities and \bar{n}_i is the surface galaxy density of each bin. Essentially, we combine the Poisson shot noise contribution with cosmic variance to obtain the total uncertainty on the power spectrum, P_κ .

Note that since this involves an integration along the line of sight, one needs to compute the matter power spectrum at multiple redshifts and interpolate between them to compute the integral in eq. (5.1). To do this, we make use of the `project_2d` module in the `Cosmosis` numerical library [67]. We compute both the linear and non-linear matter power spectrum from our simulations at $z_{\text{pk}} = \{0, 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 2.0\}$ which are then used as input to perform the integral in eq. (5.1) assuming the tomographic source redshift distribution given by eq. (5.3). Following eq. (5.1), we multiply the input power spectra that we are interpolating between by the corresponding value of $\mu^2(1 + \eta)^2/4$ at each redshift.

We compute the convergence power spectra from both linear matter power spectra and the non-linear matter power spectra from our simulations, and compare them with the Λ CDM convergence power spectrum with Euclid-like error bars. On the left hand panel of figure 9, we show the convergence power spectra as computed from the output of one of our simulations. Note that for this plot, we only consider a variation in μ , and we set η to its Λ CDM value ($\eta = 1$). We see that on large scales, i.e., small ℓ , the modified gravity spectrum is within the Λ CDM error bars, while the spectrum on non-linear scales is outside the error bars. Moreover, the linear spectrum deviates from the non-linear spectrum below $\ell = 100$.

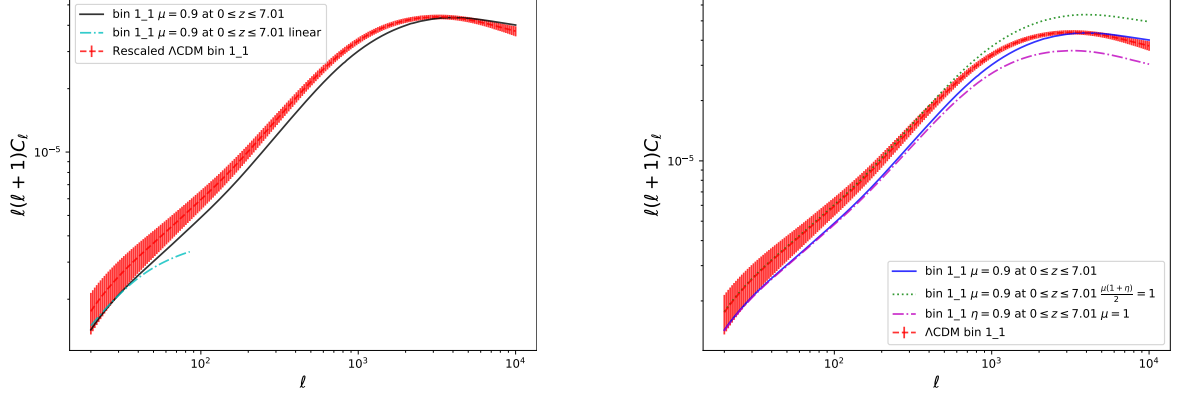


Figure 9. In the left panel, we show the weak lensing auto-power spectrum for a particular tomographic bin from Λ CDM (red-dashed) and from the simulations with $\mu = 0.9$ at $0 \leq z \leq 7.01$ (blue solid). Note that we are using the Λ CDM simulation with rescaled initial conditions at $z = 50$. In the right panel, we show three different curves with the same μ bin, but with different values of η . The first curve has $\eta = 1$ (blue solid). The second has $\eta = 0.8$ such that the prefactor $\mu(1+\eta)/2 = 1$ as in Λ CDM (green-dotted), such that the modified gravity effects only enter through the modified P_δ . Finally, we show the case where $\mu = 1$, but $\eta = 0.8$ at $0 \leq z \leq 7$, which has a different shape, i.e., different scale dependence on non-linear scales. Therefore, we show how our approach may be used by future missions to put constraints on both μ as well as η . We stress that by explicitly going to non-linear scales, we are able to break the degeneracy between η and μ that exists in linear theory.

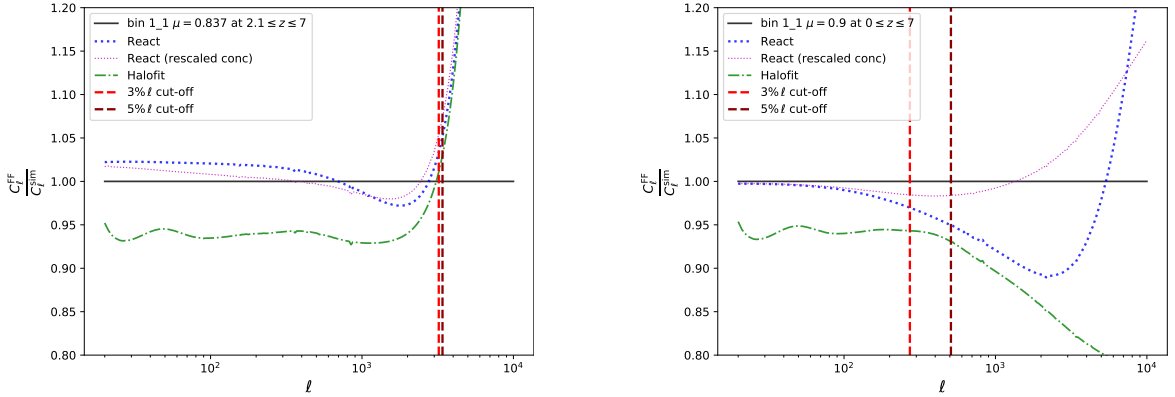


Figure 10. We show the ratio of the convergence power spectra obtained from the matter power spectra as predicted by the ReACT (blue-dotted) and halofit (green dot-dashed) formalisms with respect to those obtained from the simulation matter power spectra. The superscripts ‘FF’ and ‘sim’ imply ‘fitting function’ and ‘simulation’, respectively. We immediately see that halofit performs significantly worse than ReACT even at low ℓ , where it fails at the level of $\sim 5\%$. The vertical lines indicate the value of ℓ at which the ReACT curve deviates from the simulation curve at the level of 3% and 5%, respectively. We choose two simulations that capture the variation in the ℓ cut-offs in our sample of simulations. In the left panel is the simulation with the largest ℓ cut-offs in our sample of simulations. In the right panel, we show a simulation with one of the smallest ℓ cut-offs. The better performance of the ReACT formalism over halofit in such a model-independent analysis of modified gravity on non-linear scales is clearly demonstrated. We also show the curve with the concentration parameter modified (see figure 6) which seems to further improve the ReACT fit, see text for details.

This shows the strong increase in constraining power that can be achieved by using non-linear scales: not only are there many more scales that can be used, but also the errors are typically smaller on those scales. These results emphasise the importance of developing simulations and fitting functions for model-independent modified gravity, as examined in section 3.

We note that comparing our modified gravity parameterisation (with parameters that are only time dependent) to Λ CDM for a single observable is slightly non-trivial, because one has a free choice of which redshift to make the linear matter power spectra agree at. For our weak-lensing comparison, using the lowest redshift tomographic bin, we choose to focus on the case where the two linear spectra are equal at redshift zero (i.e. the Λ CDM simulation with rescaled initial conditions at $z = 50$). We make this choice for several reasons. It minimises the linear theory difference between the two curves, allowing us to focus on the additional information from non-linear behaviour/scales and, as long as the linear theory behaviour is similar at low redshift, the exact redshift at which they are equal is unimportant since lensing combines information from multiple redshifts. Our conclusions about the relative sizes of errors on difference scales, and the relative numbers of scales that contribute in each regime, are unchanged if a different choice is made. We will extend our analysis to a full parameter forecast for upcoming surveys in future work, which will allow this issue of parameter degeneracies to be examined in detail.

We now generalise our results to using our simulation output to probe the effects of both μ and η . As mentioned before, since the input matter power spectra from the simulations are unaffected by $\eta \neq 1$, we may model the effects of it in post-processing. In other words, the evolution of matter fluctuations is driven solely by μ , but the trajectory of photons through this distribution is affected by both η and μ . Therefore, we can effectively ignore η whilst running our N -body simulations without any loss of generality, essentially getting $P_\delta(k)$ as a function of μ only. We then use eq. (5.1) to compute the weak-lensing observables for both parameters.

As shown in eq. (5.1), the effect of varying η is captured solely by the prefactor $\mu^2(1 + \eta)^2/4$. Therefore, one would expect that a variation of η would result in an overall shift of the power spectrum. If one were to concentrate on purely linear scales, this is exactly the same as varying μ due to the fact that a time-dependent μ also simply shifts the matter power spectrum in the vertical direction(s). As a result, there is a degeneracy between time-dependent η and μ in linear theory. By explicitly going to non-linear scales, we break this degeneracy since we have already shown that a time-dependent μ introduces scale-dependent features in the matter power spectrum, and therefore in the weak-lensing convergence spectrum via P_δ in eq. (5.1), whereas η introduces no such scale dependence.⁸

The breaking of this degeneracy is clearly shown in the right panel of eq. (5.1), where we show the Λ CDM convergence spectrum with error bars, along with two different curves obtained from the same simulation, with identical $\mu(z)$, but different values of η . The green-dotted line is obtained by setting $\mu^2(1 + \eta)^2/4 = 1$ (as in Λ CDM). The solid blue line is obtained directly from the simulation, i.e., $\eta = 1$. We see that varying η while keeping μ constant leads to an overall different amplitude of the convergence spectra. We also explicitly show that if one reverts to the Λ CDM value of $\mu = 1$ while keeping the prefactor $\mu(1 + \eta)/2$ the same as in the blue line, one obtains the same shape for the power spectrum as the Λ CDM case. This can also be seen by comparing the shapes of the blue line and the green-dotted line which have identical $\mu(z)$. Thus, it is clear that the scale dependent effects from a time-

⁸In principle, η does introduce a small scale dependence due to weighting of the non-linear spectra at different times by different amounts, but this will be much smaller than the scale dependence caused by μ on non-linear scales.

dependent μ breaks the degeneracy between the two parameters that exists on linear scales, allowing the two parameters to be distinguished from each other and from their Λ CDM values. We note that the linear theory degeneracy should persist in both the auto-power spectra and the cross-power spectra. However, we leave the full behaviour of the cross-power spectra as a function of varying η and μ to future work.

These results not only show how weak-lensing observables can be constructed from our simulations for the full (μ, η) parameter space, but also further highlight the importance and gain of using non-linear scales.

5.2 Further evaluation of fitting functions

We can extend the earlier analysis of the fitting functions, by examining how well they capture the non-linear behaviour of the weak-lensing observables. We focus on the two fitting functions that performed the best in section 3.2, i.e., the **ReACT** formalism and the **halofit** fitting procedure.

In general we find that the better performance of **ReACT** compared to **halofit** is further improved when looking at the weak-lensing observables, compared to when directly examining the matter power spectrum. In particular, similarly to section 3.2, we quantify the quality of the fitting functions by calculating the ℓ_{fail} value (rather than k_{fail} as in section 3.2) at which the fitting function first fails by 3 (or 5) %. The ℓ_{fail} values at which **ReACT** fails by 3% are in the range $250 \leq \ell_{\text{fail}} \leq 4000$, however **halofit** typically fails at the 3% level even for very low ℓ . For a given fitting function, we find little correlation between the k and ℓ values at which the fitting function fails in each simulation, due to the range of k and redshift values at which the matter power spectrum contributes to the weak-lensing convolution.

To illustrate this performance, we plot the results for two representative simulations: one where **ReACT** performs well to higher ℓ , and one where it fails at much lower ℓ , approximately corresponding to the best and worst cases we studied. These simulations are shown in figure 10, where we show the ratio of the convergence power spectra calculated from the matter power spectra predicted by the fitting function and the simulation respectively. These plots show that **halofit** performs significantly worse than **ReACT**, failing at the level of $\sim 5\%$ even at low ℓ . This poor performance of **halofit** when computing weak-lensing observables is common across all simulations.

In these plots, the vertical lines show the ℓ_{fail} values, i.e., where the **ReACT** curve deviates from the simulation curve at the level of 3% and 5%, respectively. The left panel shows a simulation where the **ReACT** curve is accurate to $\ell \sim 5000$ while on the right, cut-offs are an order of magnitude smaller. These ℓ_{fail} values give an indication of the ℓ range where **ReACT** can be used to fairly reliably predict the weak-lensing observables for model-independent modified gravity studies. As described in section 3.2, we expect this performance to be improved by adjusting the concentration parameter. To demonstrate this, we include an additional curve where the concentration parameter has been modified according to the discussion at the end of section 3.2 (see figure 6 for more details). As expected, it appears that weak-lensing observables can be accurately computed to higher ℓ using this process, however we leave a detailed examination of this to future work.

These results, in combination with those in section 3.2, show that phenomenological modified gravity analyses with current data can be carried out without restricting to linear scales (e.g. [9]) or carrying out a linearisation procedure as in [68]. For these purposes, our results show that **ReACT** is the best performing fitting function (including outperforming **halofit**), particularly when using weak-lensing observables. In future work we will extend

this initial analysis to quantitatively examine a much wider range of parameter space, in preparation for analysing the data for upcoming surveys.

6 Discussion and conclusion

In this work, we have presented N -body simulations with a time-dependent strength of gravity μ , based on a framework for examining modified gravity in a model-independent way across all cosmological scales [2]. The key results of this paper are a presentation of the phenomenology of these simulations, an evaluation of existing fitting functions for capturing this phenomenology, and a demonstration of the application and importance of this framework for weak-lensing observables.

We modified the **GADGET-2** N -body code [43], and ran a series of simulations with piecewise-constant bins in redshift for μ . See section 3 for more details and table 1 for the redshift bins and the μ values in each bin.

The only fitting function calibrated from N -body simulations for the matter power spectrum on non-linear scales in phenomenological modified gravity was presented in [35]. We investigated the performance of this fitting function, as well as the Λ CDM halofit fitting procedure, the standard halo model of structure formation [see eqs. (C.2), (C.3)] and the halo model reaction [3, 60]. We compared the ratio of the modified gravity matter power spectrum to the so-called ‘pseudo’ Λ CDM power spectrum as predicted in the various formalisms to the same ratio measured from our simulations. We quantified the accuracy of each formalism by calculating the least square statistic χ^2 and k_{fail} , the wavenumber of first failure (see figure 5) for each fitting function. We found that the halo model reaction formalism performed significantly better than the others, with the notable exception being the case where the modified gravity parameters are such that the resulting power spectrum is co-incidentally very similar to the Λ CDM case. We also see qualitative evidence (see figure 6) supporting the theoretical expectation that to achieve precision in forecasting modified gravity matter power spectra, one needs to modify the Λ CDM concentration-mass relationship. We will extend the investigation presented here to a full parameter space examination and validation of the ReACT approach and the concentration issue in future work.

We present the matter power spectra from our simulations in section 4. Figures 7 and 8 show that a purely time-dependent μ induces scale-dependent features in the matter power spectrum, as reported in [35] for a simpler case. Our results also show that the shape of the power spectrum on quasi-linear as well as non-linear scales depends not only on the value of μ , but also on the redshift at which it is ‘switched on’ (i.e., different from unity) and the duration of such a modification. Most notably, we see that introducing a modification in μ at early times produces a power spectrum that either lacks ($\mu < 1$) or has excess power on the quasi-linear scales ($\mu > 1$) and vice-versa on non-linear scales, respectively. We also noticed that the peak in the power spectrum relative to Λ CDM shifts to larger scales (lower k) as one steps forward in time (from high redshift to low redshift).

To understand the physics that leads to these results, we first note that modifying μ at different epochs and for different periods of time affects the redshift at which non-linear structure starts to form. Combining this with our results indicates that a *transfer of power* occurs from smaller to larger scales or vice versa, depending on when and for how long μ is modified. This is due to the fact that at early times, the very first haloes are in the process of forming, which means that changing the rate of structure formation affects the smaller cosmological scales since those are the haloes forming at that time. Therefore, when

one returns to Λ CDM structure formation at late times, one sees a marked difference in the power spectrum. However, if one introduces a change in the rate of structure formation at late times, when galaxies have already formed, one is then affecting the rate at which the largest haloes form, via mergers. Therefore, one sees a unique change in both the quasi-linear as well as the non-linear scales of the matter power spectrum, for a given time evolution of μ .

Finally, we show the impact of these phenomenological modifications to gravity on weak-lensing observations that will be generated in future experiments like the *Euclid* satellite (see figure 9 and the discussion attached). We show that the constraining power of these experiments in the context of modified gravity is strongly dependent on the data obtained from non-linear cosmological scales, both in terms of the number of scales that are accessible and the sizes of the errors on the different scales. We further show how we can use weak-lensing observables to probe the two-parameter $\mu - \eta$ family of modified gravity models using the output of our simulations [see figure 10]. We find that by explicitly going to non-linear scales one breaks the degeneracy between these two parameters that exists on linear scales [see right panel of figure 9]. We also extend our analysis of the fitting functions to the weak-lensing context, again finding that the ReACT formalism performs the best. This shows that phenomenological modified gravity analyses with current data can be carried out without restricting to linear scales or carrying out a linearisation procedure.

The model-independent approach first elucidated in [2], and further developed here, is not a priori restricted to particular regions of model space or types of theories, and has the potential to be a powerful null test of the Λ CDM+GR paradigm. The simulations and results presented here are an important step towards realising this, and using all of the data in future surveys to put model-independent constraints on the laws of gravity that operate in the Universe.

Acknowledgments

This work was performed using the DiRAC Data Intensive service at Leicester, operated by the University of Leicester IT Services, which forms part of the STFC DiRAC HPC Facility (<https://dirac.ac.uk/>). The equipment was funded by BEIS capital funding via STFC capital grants ST/K000373/1 and ST/R002363/1 and STFC DiRAC Operations grant ST/R001014/1. DiRAC is part of the National e-Infrastructure. We also thank Lee Whitaker for valuable comments and advice. FP acknowledges the support from the grant ASI n.2018-23-HH.0.

A Convergence tests

We now present the procedure we followed to test the convergence of our N -body simulations. In the following discussion we explicitly show convergence with respect to particle number in our simulations. Using the same methods, we also tested convergence with respect to varying box-size, softening length and also different realisations of initial conditions, respectively.

In order to demonstrate convergence in our simulations, we present ratios of the matter power spectrum as this allows one to neglect realisation-dependent effects [50]. We run all our simulations in a cosmological box of side $250 \text{ Mpc } h^{-1}$. In the left panel of figure 11, we compute the ratio of the matter spectrum from Λ CDM simulations with respect to our highest resolution reference simulation, increasing the particle in steps. The green line denotes the k value at which the second-highest resolution simulation (512^3 particles) disagrees with our

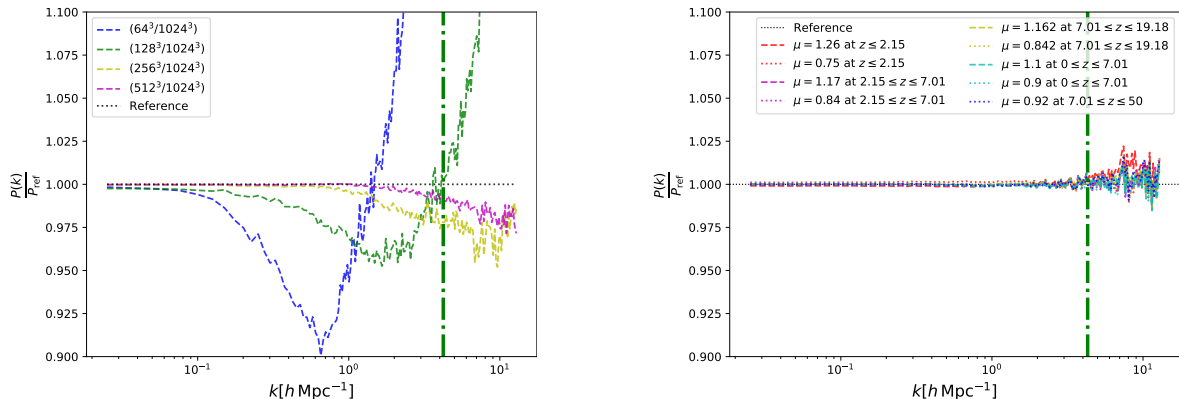


Figure 11. *Left panel:* the ratio of the power spectra as a function of scale, with each line labelled by the particle number of the simulation they represent. All simulations have a box size of $250 \text{ Mpc } h^{-1}$ comoving. The green vertical line represents the wavelength at which the ratio of the power spectra deviates at the level of 1% from the simulations with the best resolution. This panel shows the level of convergence we obtain in our Λ CDM simulations. *Right panel:* a similar procedure was carried out for the modified gravity simulations, but with a key difference. For each simulation, we calculate the ratio $S(k)$ of the power spectrum from the simulation to the power spectrum from the Λ CDM simulation. We then plot the ratio of $S(k)$ computed from simulations with 512^3 particles to simulations with 1024^3 particles. The green vertical line is plotted at the identical point as before. Therefore, we show that our modified gravity simulations have similar convergence both with respect to each other as well as with respect to our Λ CDM simulations. Note that on linear scales, all the ratios are within 0.5% of 1, which justifies our choice of μ in each bin. For our production simulations, we use a particle number of 1024^3 (with an equally fine mesh) and a box size of $250 \text{ Mpc } h^{-1}$.

highest resolution (1024^3 particles) simulation, which is an indicator of the scale up to which our results are reliable.

We now show that we have achieved the same level of convergence in our modified gravity simulations. In the right panel, we show the convergence of our modified gravity simulations by plotting a ratio of ratios. We evaluate the ratio of matter power spectra from the simulation with 512^3 particles with the simulation with 1024^3 particles in both our Λ CDM simulations as well as our modified gravity simulations. We then plot S_{512}/S_{1024} , and show that they agree up to the same wavenumber as the left-hand-side panel, again shown by the green line. The fact that the vertical green line is at an identical wavenumber on both our panels demonstrates that we have achieved convergence for both our Λ CDM simulations as well as our modified gravity simulations.

B Binning in μ

We now describe the procedure we adopted to bin μ in redshift. We reiterate this parameter space is multi-dimensional, which simultaneously covers a large portion of the parameter space while isolating non-linear behaviour is a difficult task. We focus on the latter in this work, as our intention is to understand the phenomenology.

As mentioned in section 3, we divide the redshift spanned by our simulations into 2 and 4 bins of equal Λ CDM growth. In other words, in the case where we have 2 redshift bins for μ , we have that $D^{\Lambda\text{CDM}}(z_1)/D^{\Lambda\text{CDM}}(z_2) = D^{\Lambda\text{CDM}}(z_2)/D^{\Lambda\text{CDM}}(z_3)$. Setting $z_1 = 0$ and $z_3 = 50$, we obtain

$$D(z_2)^2 = D(z_1)D(z_3), \quad (\text{B.1})$$

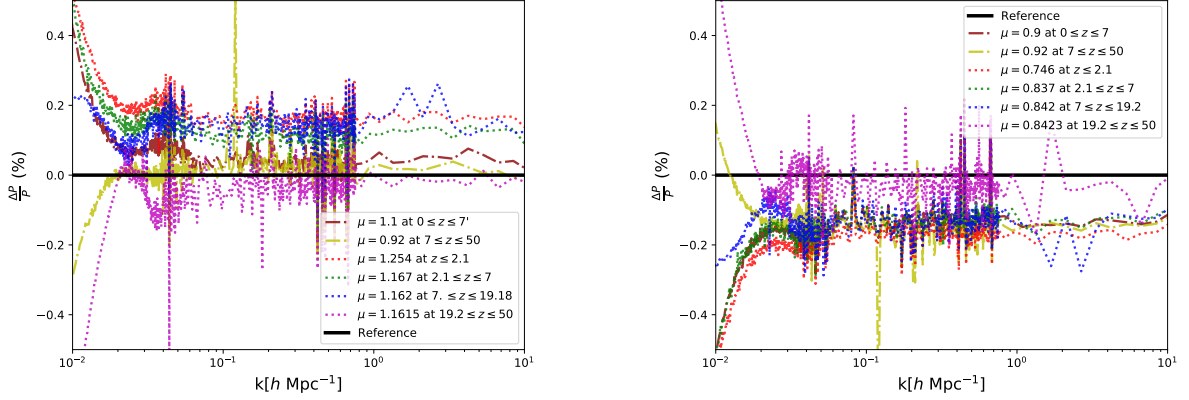


Figure 12. The relative difference in the linear power spectra at redshift zero to the reference case (chosen to be the case where μ is constant) from runs of our modified CLASS code.

from which we can compute $z_2 \approx 7.01$. We follow a similar procedure in the case where we have 4 bins in redshift. We consider the bin $0 \leq z \leq 7.01$ as our reference in the context of choosing the value of μ . We set $\mu = 1.1$ and 0.9 , in this bin, respectively, and compute $D(0)$. For all the other bins, we choose μ such that the growth factor at $z = 0$ is identical to the value we computed for our reference bin. In this way, we obtain the μ values in table 1. In order to confirm that these values of μ indeed correspond to identical growth factors at $z = 0$, we modified the equations of motion solved by CLASS for a time-dependent μ as outlined in [69]. In figure 12 we show that the relative differences in the linear power spectra for all the cases in table 1 are within 0.5%, in accordance with the right panel in figure 11, which shows that our modified CLASS code performs as expected. Therefore, we run two sets of seven simulations in which $\mu \neq 1$ in one, two or four redshift bins, with the matter power spectrum measured from all the simulations being equal on linear scales at $z = 0$.

We now describe the implementation of the μ smoothing in CLASS as well as GADGET-2. We assume that the input to the code includes an array of values for μ as well as an array of redshifts that denote the bins. We then use a combination of two error functions given by

$$E_1 = C_1 \text{erf}((z_{\text{array}}[i] - z)/X) + C_2, \quad (\text{B.2})$$

$$E_2 = C_1 \text{erf}(-(z_{\text{array}}[i+1] - z)/X) + C_2, \quad (\text{B.3})$$

where C_1 , C_2 and X are parameters that determine the amplitude, midpoint and width of the error function, respectively, to determine μ at each redshift. The iterative variable i is incremented once the current redshift (associated to the timestep in the code) is lower than the midpoint of the current bin, i.e., when $z < (z_{\text{array}}[i+1] + z_{\text{array}}[i])/2$. The constants C_1 and C_2 are given by

$$C_1 = \mu_{\text{array}}[i+1] - \mu_{\text{array}}[i]/2, \quad (\text{B.4})$$

$$C_2 = \mu_{\text{array}}[i+1] + \mu_{\text{array}}[i]/2, \quad (\text{B.5})$$

The value of μ is given by

$$\mu = E_1 \text{ if } z > z_{\text{array}}[i+1] - z[i]/2, \quad (\text{B.6})$$

$$\mu = E_2 \text{ if } z < z_{\text{array}}[i+1] - z[i]/2. \quad (\text{B.7})$$

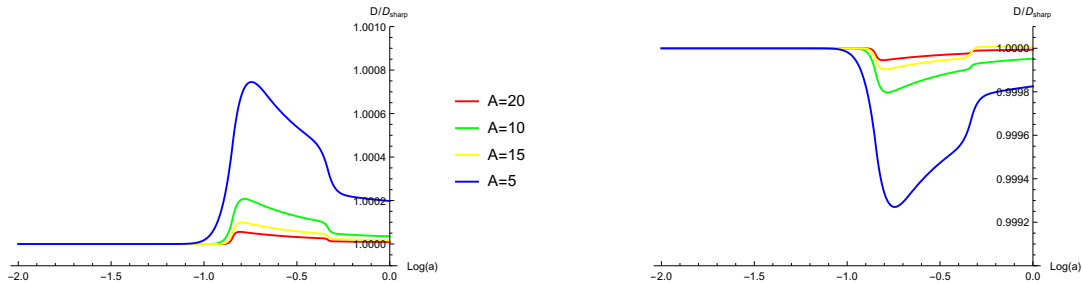


Figure 13. The plot of the growth factor as a function of redshift in the case where $\mu > 1$ (left) and $\mu < 1$ (right) at $2.15 \leq z \leq 7.01$.

We have chosen the width function $X = (1/A) \ln(\Delta z_{\text{bin}})$ with A being a tolerance parameter that determines the width of the transition and Δz_{bin} is the bin-width. This choice ensures that the width of the transition naturally becomes shorter at lower redshifts where the growth factor is more sensitive to redshift (where the timestep in the code is small).

To estimate the impact of the smoothing, we have carried out two tests. The first test is the ratio of the power spectrum with the smoothed μ to the un-smooth case for different values of A . In this case, we have averaged the ratio over all k values to obtain a “goodness of fit” parameter, which ideally should be equal to zero. The second test was to plot the growth factor as a function of redshift, specifically over the range of redshifts where the smoothing is applied. The results from both these tests in **CLASS** are shown in figure 13. On implementing this test in **GADGET-2**, we found that over the range of values we considered for the smoothing parameter, the differences between the smoothed and un-smoothed case were smaller than 0.1%.

C Fitting functions

We now present an overview of the different fitting functions we examined in section 3.2. Note that we also examine the performance of the standard **halofit** procedure [56–58] which we do not describe in detail here. Not that while this is not a detailed examination of these fitting functions, we present the procedure that we followed to actually predict $R(k)$ in order to compare with our simulations.

C.1 Cui et al.

The fitting function derived in [35] is given by

$$R(x, \mu) = \exp \left\{ ((1 - \mu)B(\mu)x^{C(\mu)}) \right\}, \quad (\text{C.1})$$

where $R(x, \mu)$ as mentioned in the main text is the ratio of the matter power spectrum in modified gravity relative to the Λ CDM pseudo spectrum (with equal linear growth). $B(\mu) = 0.0429 + 0.133\mu^{-4}$, $C(\mu) = 0.573$ and $x = \Delta^2(k, z, \mu = 1)$ is the dimensionless power spectrum $P(k)k^3/(2\pi)$. The main feature of this function is that it predicts $R > 1$ when $\mu < 1$ and vice-versa when $\mu > 1$. In other words, this fitting function assumes that the non-linear evolution is slower when the growth rate is increased due to μ being larger than its value in Λ CDM, and vice versa. While we have observed a similar behaviour in some of our simulations, we find that this is not true in the general case.

We demonstrate in figure 14 that one can reproduce with reasonable success the matter power spectrum in the case where μ is constant throughout the simulation time period,

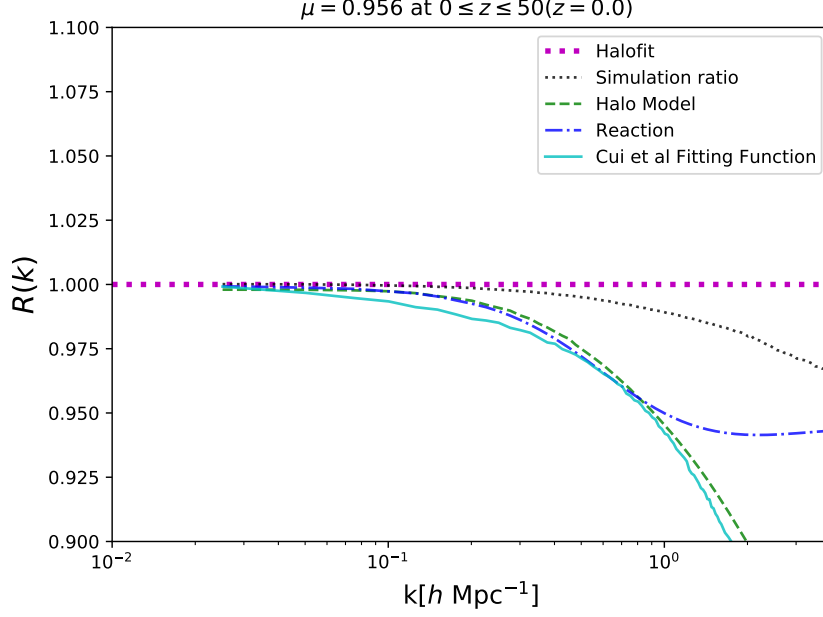


Figure 14. The ratio of the matter power spectrum in the simulation with $\mu = 0.956$ throughout the simulation period. The Cui et al. fitting function [35] is designed to model the power spectrum in this case, where μ is constant.

although even in this case we see that ReACT outperforms the other fitting functions that we consider.

C.2 Halo model

The standard halo model consists of the 1-halo term, given by

$$P^{1h}(k) = \int dm n(m) \left(\frac{m}{\bar{\rho}} \right)^2 |u(k|m, z)|^2, \quad (C.2)$$

which is the contribution to the power spectrum from a single halo, along with the 2-halo term given by

$$P^{2h}(k) = \int dm_1 n(m_1) \left(\frac{m_1}{\bar{\rho}} \right) |u(k|m_1, z)| \int dm_2 n(m_2) \left(\frac{m_2}{\bar{\rho}} \right) |u(k|m_2, z)| P_{hh}(k, z|m_1, m_2), \quad (C.3)$$

which is the contribution from 2 different haloes, where $n(m)$ is the mass function, $u(k, z|m)$ is the Fourier transform of the density profile dark matter haloes of mass m at redshift z and P_{hh} is the halo power spectrum, related to the two-point-correlation function of the haloes themselves and quantifies the interaction between two haloes. Of these, the first term dominates on small non-linear scales, i.e., at large k , while the 2-halo term dominates on large scales. Note that the 2-halo term is usually well approximated by the linear matter powers spectrum in the literature.

In order to compute the 1-halo term, we calculate the mass function and the bias term assuming a time-dependent μ according to the Sheth-Tormen procedure [54, 55] and perform the integrals in eqs. (C.2)(C.3). We then obtain the full non-linear matter power spectrum by summing up the 1-halo term and 2-halo term. As a cross-check we ensured that our results are replicated by summing the 1-halo term and the linear matter power spectrum.

Increasing the value of μ increases the rate of structure formation which means that haloes form earlier, resulting in smaller concentration parameters associated to the haloes. Since the Fourier transform of the density profiles is directly proportional to the concentration parameter, the contribution from the 1-halo term is smaller, resulting in a lack of power on smaller scales [35] (with the opposite taking place for $\mu < 1$).

C.3 Halo model reaction

The halo model reaction formalism integrates a modified halo model and the spherical collapse model together to predict the matter power spectrum on non-linear scales. As mentioned before, the halo model suffers from a lack of accuracy on quasi-linear scales as in the original recipe. The authors go one step further to address this problem via the reaction term, given by

$$\mathcal{R}(a, k) = \frac{\left[(1 - \epsilon(z))e^{-k/k_*(z)} + \epsilon(z) \right] P_{\text{mg}}^{\text{lin}}(z, k) + P_{\text{mg}}^{\text{1h}}(z, k)}{P_{\text{pseudo}}^{\text{hm}}(z, k)}, \quad (\text{C.4})$$

where the subscript ‘pseudo’ represents the pseudo spectrum and k_* is computed from perturbation theory [70]. Therefore, the quantity \mathcal{R} , the so-called reaction, is a ratio of halo-model power spectra. In order to obtain a prediction of the modified gravity power spectrum, one simply multiplies \mathcal{R} by the full non-linear Λ CDM pseudo spectrum. To understand how well ReACT captures our simulations, we use the ReACT code to calculate $R(a, k) = P_{\text{mg}}/P_{\Lambda\text{CDM}}^{\text{pseudo}} = \mathcal{R}$.

The quantity \mathcal{R} has the following basic features:

- On linear scales, $\mathcal{R} \rightarrow 1$ by construction since the pseudo-spectrum has the same linear growth as the modified gravity spectrum.
- On small non-linear scales the reaction is given by the ratio of the 1-halo terms, i.e., $\lim_{k \rightarrow k_{\text{nl}}} \mathcal{R} \rightarrow P_{\text{mg}}^{\text{1h}}(z, k)/P_{\text{pseudo}}^{\text{1h}}(z, k)$.

The reaction term in eq. (3.1) may be computed from the following equations

$$P_{\text{pseudo}}^{\text{hm}} = P_{\text{pseudo}}^{\text{lin}} + P_{\text{pseudo}}^{\text{1h}}, \quad (\text{C.5})$$

$$\epsilon = \lim_{k \rightarrow 0} \frac{P_{\text{mg}}^{\text{1h}}(z, k)}{P_{\text{pseudo}}^{\text{1h}}(z, k)}, \quad (\text{C.6})$$

$$k_*(z) = -\bar{k} \left(\ln \left[\frac{A(\bar{k}, z)}{P_{\text{mg}}^{\text{lin}}(z, k)} - \epsilon(z) - \ln(1 - \epsilon(z)) \right] \right), \quad (\text{C.7})$$

$$A(k, z) = \frac{P_{\text{mg}}^{\text{1-loop}}(z, k) + P_{\text{mg}}^{\text{1h}}(z, k)}{P_{\text{pseudo}}^{\text{1-loop}}(z, k) + P_{\text{pseudo}}^{\text{1h}}(z, k)} P_{\text{pseudo}}^{\text{hm}}(z, k) - P_{\text{mg}}^{\text{1h}}(z, k), \quad (\text{C.8})$$

where the 1-loop terms may be computed from Fourier transforming the Poisson equation [3, 60]

$$-k^2 \phi(a, k) = \frac{3\Omega_{\text{m}}(a)aH(a)}{2} \mu(a, k) \delta(a, k) + \tilde{S}(a, k), \quad (\text{C.9})$$

where the source function $\tilde{S}(a, k)$ encodes non-linear corrections to the Newtonian Poisson equation that appear at higher order. Essentially, one can derive $\tilde{S}(a, k)$ for specific models and the corresponding screening mechanisms that operate on the relevant scales. The authors in [60] calculate $\tilde{S}(a, k)$ for $f(R)$ and DGP gravity in order to reproduce the full

non-linear matter power spectrum in modified gravity. Note that the **ReACT** code is set up to automatically calculate $\tilde{S}(a, k)$ for these two models and Λ CDM (where it is zero).

In contrast to [60], our work is designed to understand the phenomenology while remaining model-agnostic and therefore, we neglect any screening mechanisms and indeed any scale dependent corrections to the Poisson equation given in eq. (2.6). Therefore, in order to reproduce the matter power spectrum in our formalism within the **ReACT** code, we only modify the μ parameter and set $\tilde{S}(a, k) = 0$ and indeed all the other modified gravity parameters within the code to their respective values in Λ CDM. This simplifies the numerical implementation of the **ReACT** while still being able to estimate the non-linear effects of modifying μ .

References

- [1] C. Garcia-Quintero, M. Ishak and O. Ning, *Current constraints on deviations from general relativity using binning in redshift and scale*, *JCAP* **12** (2020) 018 [[arXiv:2010.12519](#)] [[INSPIRE](#)].
- [2] D.B. Thomas, *Cosmological gravity on all scales: simple equations, required conditions, and a framework for modified gravity*, *Phys. Rev. D* **101** (2020) 123517 [[arXiv:2004.13051](#)] [[INSPIRE](#)].
- [3] M. Cataneo et al., *On the road to percent accuracy: non-linear reaction of the matter power spectrum to dark energy and modified gravity*, *Mon. Not. Roy. Astron. Soc.* **488** (2019) 2121 [[arXiv:1812.05594](#)] [[INSPIRE](#)].
- [4] S. Weinberg, *Gravitation and cosmology: principles and applications of the general theory of relativity*, John Wiley & Sons, U.S.A. (1972).
- [5] M. Schumann, *Direct detection of WIMP dark matter: concepts and status*, *J. Phys. G* **46** (2019) 103003.
- [6] I.G. Irastorza and J. Redondo, *New experimental approaches in the search for axion-like particles*, *Prog. Part. Nucl. Phys.* **102** (2018) 89 [[arXiv:1801.08127](#)] [[INSPIRE](#)].
- [7] T. Clifton, P.G. Ferreira, A. Padilla and C. Skordis, *Modified gravity and cosmology*, *Phys. Rept.* **513** (2012) 1 [[arXiv:1106.2476](#)] [[INSPIRE](#)].
- [8] K. Koyama, *Cosmological tests of modified gravity*, *Rept. Prog. Phys.* **79** (2016) 046902 [[arXiv:1504.04623](#)] [[INSPIRE](#)].
- [9] R.A. Battye and J.A. Pearson, *Effective action approach to cosmological perturbations in dark energy and modified gravity*, *JCAP* **07** (2012) 019 [[arXiv:1203.0398](#)] [[INSPIRE](#)].
- [10] G. Gubitosi, F. Piazza and F. Vernizzi, *The effective field theory of dark energy*, *JCAP* **02** (2013) 032 [[arXiv:1210.0201](#)] [[INSPIRE](#)].
- [11] P.G. Ferreira, T. Baker and C. Skordis, *Testing general relativity with cosmology: a synopsis of the parametrized post-Friedmann approach*, *Gen. Rel. Grav.* **46** (2014) 1788 [[INSPIRE](#)].
- [12] J.N. Dossett, M. Ishak and J. Moldenhauer, *Testing general relativity at cosmological scales: implementation and parameter correlations*, *Phys. Rev. D* **84** (2011) 123001 [[arXiv:1109.4583](#)] [[INSPIRE](#)].
- [13] E. Bertschinger, *On the growth of perturbations as a test of dark energy*, *Astrophys. J.* **648** (2006) 797 [[astro-ph/0604485](#)] [[INSPIRE](#)].
- [14] L. Pogosian, A. Silvestri, K. Koyama and G.-B. Zhao, *How to optimally parametrize deviations from general relativity in the evolution of cosmological perturbations?*, *Phys. Rev. D* **81** (2010) 104023 [[arXiv:1002.2382](#)] [[INSPIRE](#)].

- [15] SUPERNOVA SEARCH TEAM collaboration, *Observational evidence from supernovae for an accelerating universe and a cosmological constant*, *Astron. J.* **116** (1998) 1009 [[astro-ph/9805201](#)] [[INSPIRE](#)].
- [16] SUPERNOVA COSMOLOGY PROJECT collaboration, *Measurements of Ω and Λ from 42 high redshift supernovae*, *Astrophys. J.* **517** (1999) 565 [[astro-ph/9812133](#)] [[INSPIRE](#)].
- [17] A.G. Riess et al., *New Hubble space telescope discoveries of type Ia supernovae at $z \gtrsim 1$: narrowing constraints on the early behavior of dark energy*, *Astrophys. J.* **659** (2007) 98 [[astro-ph/0611572](#)] [[INSPIRE](#)].
- [18] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [[arXiv:1807.06209](#)] [[INSPIRE](#)].
- [19] PLANCK collaboration, *Planck 2018 results. VIII. Gravitational lensing*, *Astron. Astrophys.* **641** (2020) A8 [[arXiv:1807.06210](#)] [[INSPIRE](#)].
- [20] DES collaboration, *Dark Energy Survey year 1 results: cosmological constraints from galaxy clustering and weak lensing*, *Phys. Rev. D* **98** (2018) 043526 [[arXiv:1708.01530](#)] [[INSPIRE](#)].
- [21] DES collaboration, *Density split statistics: cosmological constraints from counts and lensing in cells in DES Y1 and SDSS data*, *Phys. Rev. D* **98** (2018) 023507 [[arXiv:1710.05045](#)] [[INSPIRE](#)].
- [22] L. Lombriser, *A parametrisation of modified gravity on nonlinear cosmological scales*, *JCAP* **11** (2016) 039 [[arXiv:1608.00522](#)] [[INSPIRE](#)].
- [23] T. Clifton, C.S. Gallagher, S. Goldberg and K.A. Malik, *Viable gauge choices in cosmologies with nonlinear structures*, *Phys. Rev. D* **101** (2020) 063530 [[arXiv:2001.00394](#)] [[INSPIRE](#)].
- [24] H.A. Buchdahl, *Non-linear Lagrangians and cosmological theory*, *Mon. Not. Roy. Astron. Soc.* **150** (1970) 1.
- [25] G.R. Dvali, G. Gabadadze and M. Porrati, *4D gravity on a brane in 5D Minkowski space*, *Phys. Lett. B* **485** (2000) 208 [[hep-th/0005016](#)] [[INSPIRE](#)].
- [26] H.A. Winther et al., *Modified gravity N-body code comparison project*, *Mon. Not. Roy. Astron. Soc.* **454** (2015) 4208 [[arXiv:1506.06384](#)] [[INSPIRE](#)].
- [27] H.A. Winther, K. Koyama, M. Manera, B.S. Wright and G.-B. Zhao, *COLA with scale-dependent growth: applications to screened modified gravity models*, *JCAP* **08** (2017) 006 [[arXiv:1703.00879](#)] [[INSPIRE](#)].
- [28] I. Milillo, D. Bertacca, M. Bruni and A. Maselli, *Missing link: a nonlinear post-Friedmann framework for small and large scales*, *Phys. Rev. D* **92** (2015) 023519 [[arXiv:1502.02985](#)] [[INSPIRE](#)].
- [29] C.M. Will, *Theory and experiment in gravitational physics*, Cambridge University Press, Cambridge, U.K. (1993).
- [30] D.B. Thomas, M. Bruni, K. Koyama, B. Li and G.-B. Zhao, *$f(R)$ gravity on non-linear scales: the post-Friedmann expansion and the vector potential*, *JCAP* **07** (2015) 051 [[arXiv:1503.07204](#)] [[INSPIRE](#)].
- [31] F. Hassani and L. Lombriser, *N-body simulations for parametrized modified gravity*, *Mon. Not. Roy. Astron. Soc.* **497** (2020) 1885 [[arXiv:2003.05927](#)] [[INSPIRE](#)].
- [32] T. Clifton and V.A.A. Sanghai, *Parametrizing theories of gravity on large and small scales in cosmology*, *Phys. Rev. Lett.* **122** (2019) 011301 [[arXiv:1803.01157](#)] [[INSPIRE](#)].
- [33] M. Kopp, C. Skordis, D.B. Thomas and S. Ilić, *Dark matter equation of state through cosmic history*, *Phys. Rev. Lett.* **120** (2018) 221102 [[arXiv:1802.09541](#)] [[INSPIRE](#)].
- [34] F. Pace, R.A. Battye, B. Bolliet and D. Trinh, *Dark sector evolution in Horndeski models*, *JCAP* **09** (2019) 018 [[arXiv:1905.06795](#)] [[INSPIRE](#)].

- [35] W. Cui, P. Zhang and X. Yang, *Nonlinearities in modified gravity cosmology. I. Signatures of modified gravity in the nonlinear matter power spectrum*, *Phys. Rev. D* **81** (2010) 103528 [[arXiv:1001.5184](#)] [[INSPIRE](#)].
- [36] Y. Zhang, P. Zhang, X. Yang and W. Cui, *Nonlinearities in modified gravity cosmology. II. Impacts of modified gravity on the halo properties*, *Phys. Rev. D* **87** (2013) 023521 [[arXiv:1301.3255](#)] [[INSPIRE](#)].
- [37] H.F. Stabenau and B. Jain, *N-body simulations of alternate gravity models*, *Phys. Rev. D* **74** (2006) 084007 [[astro-ph/0604038](#)] [[INSPIRE](#)].
- [38] I. Laszlo and R. Bean, *Nonlinear growth in modified gravity theories of dark energy*, *Phys. Rev. D* **77** (2008) 024048 [[arXiv:0709.0307](#)] [[INSPIRE](#)].
- [39] D.B. Thomas and C.R. Contaldi, *Viability of the cluster mass function formalism in parametrised modified gravity*, [arXiv:1112.6378](#) [[INSPIRE](#)].
- [40] J. Zhang, R. An, S. Liao, W. Luo, Z. Li and B. Wang, *Fully self-consistent cosmological simulation pipeline for interacting dark energy models*, *Phys. Rev. D* **98** (2018) 103530 [[arXiv:1811.01519](#)] [[INSPIRE](#)].
- [41] M. Baldi, *Clarifying the effects of interacting dark energy on linear and nonlinear structure formation processes*, *Mon. Not. Roy. Astron. Soc.* **414** (2011) 116 [[arXiv:1012.0002](#)] [[INSPIRE](#)].
- [42] F. Hassani, J. Adamek, M. Kunz and F. Vernizzi, *k-evolution: a relativistic N-body code for clustering dark energy*, *JCAP* **12** (2019) 011 [[arXiv:1910.01104](#)] [[INSPIRE](#)].
- [43] V. Springel, *The cosmological simulation code GADGET-2*, *Mon. Not. Roy. Astron. Soc.* **364** (2005) 1105 [[astro-ph/0505010](#)] [[INSPIRE](#)].
- [44] V. Springel et al., *Simulating the joint evolution of quasars, galaxies and their large-scale distribution*, *Nature* **435** (2005) 629 [[astro-ph/0504097](#)] [[INSPIRE](#)].
- [45] M. Boylan-Kolchin, V. Springel, S.D.M. White, A. Jenkins and G. Lemson, *Resolving cosmic structure formation with the Millennium-II simulation*, *Mon. Not. Roy. Astron. Soc.* **398** (2009) 1150 [[arXiv:0903.3041](#)] [[INSPIRE](#)].
- [46] R.E. Angulo, V. Springel, S.D.M. White, A. Jenkins, C.M. Baugh and C.S. Frenk, *Scaling relations for galaxy clusters in the Millennium-XXL simulation*, *Mon. Not. Roy. Astron. Soc.* **426** (2012) 2046 [[arXiv:1203.3216](#)] [[INSPIRE](#)].
- [47] D.B. Thomas and C.R. Contaldi, *Testing model independent modified gravity with future large scale surveys*, *JCAP* **12** (2011) 013 [[arXiv:1107.0727](#)] [[INSPIRE](#)].
- [48] A. Jenkins, *Second-order Lagrangian perturbation theory initial conditions for resimulations*, *Mon. Not. Roy. Astron. Soc.* **403** (2010) 1859.
- [49] M. Crocce, S. Pueblas and R. Scoccimarro, *Transients from initial conditions in cosmological simulations*, *Mon. Not. Roy. Astron. Soc.* **373** (2006) 369 [[astro-ph/0606505](#)] [[INSPIRE](#)].
- [50] P. McDonald, H. Trac and C. Contaldi, *Dependence of the non-linear mass power spectrum on the equation of state of dark energy*, *Mon. Not. Roy. Astron. Soc.* **366** (2006) 547 [[astro-ph/0505565](#)] [[INSPIRE](#)].
- [51] J. Lesgourgues, *The Cosmic Linear Anisotropy Solving System (CLASS) I: overview*, [arXiv:1104.2932](#) [[INSPIRE](#)].
- [52] S. Colombi and D. Novikov, *POWMES: measuring the power spectrum in an N-body simulation*, [ascl:1110.017](#), October 2011.
- [53] S. Colombi, A.H. Jaffe, D. Novikov and C. Pichon, *Accurate estimators of power spectra in N-body simulations*, *Mon. Not. Roy. Astron. Soc.* **393** (2009) 511 [[arXiv:0811.0313](#)] [[INSPIRE](#)].

- [54] R.K. Sheth and G. Tormen, *Large scale bias and the peak background split*, *Mon. Not. Roy. Astron. Soc.* **308** (1999) 119 [[astro-ph/9901122](#)] [[INSPIRE](#)].
- [55] R.K. Sheth and G. Tormen, *An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier*, *Mon. Not. Roy. Astron. Soc.* **329** (2002) 61 [[astro-ph/0105113](#)] [[INSPIRE](#)].
- [56] R.E. Smith et al., *Stable clustering, the halo model and non-linear cosmological power spectra*, *Mon. Not. Roy. Astron. Soc.* **341** (2003) 1311.
- [57] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya and M. Oguri, *Revising the halofit model for the nonlinear matter power spectrum*, *Astrophys. J.* **761** (2012) 152 [[arXiv:1208.2701](#)] [[INSPIRE](#)].
- [58] S. Bird, M. Viel and M.G. Haehnelt, *Massive neutrinos and the non-linear matter power spectrum*, *Mon. Not. Roy. Astron. Soc.* **420** (2012) 2551.
- [59] A. Cooray and R.K. Sheth, *Halo models of large scale structure*, *Phys. Rept.* **372** (2002) 1 [[astro-ph/0206508](#)] [[INSPIRE](#)].
- [60] B. Bose, M. Cataneo, T. Tröster, Q. Xia, C. Heymans and L. Lombriser, *On the road to per cent accuracy IV: ReACT — computing the non-linear power spectrum beyond Λ CDM*, *Mon. Not. Roy. Astron. Soc.* **498** (2020) 4650 [[arXiv:2005.12184](#)] [[INSPIRE](#)].
- [61] I. Smail, R.S. Ellis and M.J. Fitchett, *Gravitational lensing of distant field galaxies by rich clusters: I — faint galaxy redshift distributions*, *Mon. Not. Roy. Astron. Soc.* **270** (1994) 245 [[astro-ph/9402048](#)] [[INSPIRE](#)].
- [62] W. Hu, *Power spectrum tomography with weak lensing*, *Astrophys. J. Lett.* **522** (1999) L21 [[astro-ph/9904153](#)] [[INSPIRE](#)].
- [63] P.G. Castro, A.F. Heavens and T.D. Kitching, *Weak lensing analysis in three dimensions*, *Phys. Rev. D* **72** (2005) 023516 [[astro-ph/0503479](#)] [[INSPIRE](#)].
- [64] S. Casas, M. Kunz, M. Martinelli and V. Pettorino, *Linear and non-linear modified gravity forecasts with future surveys*, *Phys. Dark Univ.* **18** (2017) 73 [[arXiv:1703.01271](#)] [[INSPIRE](#)].
- [65] A. Spurio Mancini, R. Reischke, V. Pettorino, B.M. Schäfer and M. Zumalacárregui, *Testing (modified) gravity with 3D and tomographic cosmic shear*, *Mon. Not. Roy. Astron. Soc.* **480** (2018) 3725 [[arXiv:1801.04251](#)] [[INSPIRE](#)].
- [66] EUCLID collaboration, *Euclid preparation: VII. Forecast validation for Euclid cosmological probes*, *Astron. Astrophys.* **642** (2020) A191 [[arXiv:1910.09273](#)] [[INSPIRE](#)].
- [67] J. Zuntz et al., *CosmoSIS: modular cosmological parameter estimation*, *Astron. Comput.* **12** (2015) 45 [[arXiv:1409.3409](#)] [[INSPIRE](#)].
- [68] DES collaboration, *Dark Energy Survey year 1 results: constraints on extended cosmological models from galaxy clustering and weak lensing*, *Phys. Rev. D* **99** (2019) 123505 [[arXiv:1810.02499](#)] [[INSPIRE](#)].
- [69] A. Zucca, L. Pogosian, A. Silvestri and G.-B. Zhao, *MGCAMB with massive neutrinos and dynamical dark energy*, *JCAP* **05** (2019) 001 [[arXiv:1901.05956](#)] [[INSPIRE](#)].
- [70] F. Bernardeau, S. Colombi, E. Gaztanaga and R. Scoccimarro, *Large scale structure of the universe and cosmological perturbation theory*, *Phys. Rept.* **367** (2002) 1 [[astro-ph/0112551](#)] [[INSPIRE](#)].